PUBLICATION I

Darzacq X., Jády B.E., Verheggen C., Kiss A. M., Bertrand E. and Kiss T.

CAJAL BODY-SPECIFIC SMALL NUCLEAR RNAS: A NOVEL CLASS OF 2'-O-METHYLATION AND PSEUDOURIDYLATION GUIDE RNAS.

EMBO J., **21**, 2746-2756, 2002

INTRODUCTION

Eukaryotic ribosomal RNAs and spliceosomal snRNAs possess numerous post-transcriptionaly synthesised 2'-*O*-ribose-methylated nucleotides and pseudouridines which are important to the correct function of ribosome and spliceosome. 2'-*O*-methylation and pseudouridylation of ribosomal RNAs and the RNAPIII-transcribed U6 spliceosomal snRNA occur in the nucleolus and are directed by box C/D and box H/ACA modification guide RNAs, respectively. By screening a cDNA library of human snRNAs we have identified eight novel modification guide RNAs which were predicted to direct 2'-*O*-methylation and pseudouridylation of the RNAPII-transcribed U1, U2, U4 and U5 spliceosomal snRNAs. We have demonstrated that these novel guide RNAs, which indeed function in modification of Sm snRNAs, specifically accumulate in the Cajal bodies. Therefore, the new guide RNAs were called small Cajal body-specific RNAs (scaRNAs). The Cajal body is an evolutionarily conserved nucleoplasmic organelle, the function of which has remained highly elusive for more than a century. Our results strongly indicate that at least one function of the Cajal body is in post-transcriptional modification of spliceosomal snRNAs

The EMBO Journal Vol.21 No.11 pp.2746–2756, 2002

Cajal body-specific small nuclear RNAs: a novel class of 2′-*O*-methylation and pseudouridylation guide RNAs

Xavier Darzacq, Beáta E.Jády[1], Céline Verheggen[2], Arnold M.Kiss[1], Edouard Bertrand[2] and Tamás Kiss[3]

Laboratoire de Biologie Moléculaire Eucaryote du CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, [2]Institut de Génétique Moléculaire, Montpellier, France and [1]Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary

[3]Corresponding author
e-mail: tamas@ibcg.biotoul.fr

Cajal (coiled) bodies are conserved subnuclear organelles that are present in the nucleoplasm of both animal and plant cells. Although Cajal bodies were first described nearly 100 years ago, their function has remained largely speculative. Here, we describe a novel class of human small nuclear RNAs that localize specifically to Cajal bodies. The small Cajal body-specific RNAs (scaRNAs) are predicted or have already been demonstrated to function as guide RNAs in site-specific synthesis of 2′-*O*-ribose-methylated nucleotides and pseudouridines in the RNA polymerase II-transcribed U1, U2, U4 and U5 spliceosomal small nuclear RNAs (snRNAs). Our results provide strong support for the idea that the Cajal body, this mysterious nuclear organelle, provides the cellular locale for post-transcriptional modification of spliceosomal snRNAs.

Keywords: Cajal body/RNA modification/scaRNA/spliceosomal snRNA/snoRNA

Introduction

In eukaryotic cells, biogenesis of cytoplasmic mRNAs, tRNAs, rRNAs, small nuclear RNAs (snRNAs) and nucleolar RNAs (snoRNAs) takes place in the nucleus. The interphase nucleus contains at least a dozen morphologically distinct subdomains, called nuclear bodies, which are believed to function in the synthesis, processing, modification or transport of various cellular RNAs. However, with the exception of the nucleolus, which has long been known as the site of the synthesis and maturation of cytoplasmic rRNAs, the precise function of nuclear bodies remained largely elusive (Lamond and Earnshaw, 1998; Matera, 1999; Lewis and Tollervey, 2000). Besides the nucleolus, the Cajal (coiled) body is the most prominent and most extensively studied nucleoplasmic organelle (Matera, 1998; Gall, 2000). Cajal bodies are enriched with spliceosomal and nucleolar ribonucleoproteins (snRNPs and snoRNPs) and components of the basal transcription machinery. But, despite the rapidly growing list of its molecular components, the precise function of the Cajal body is still unknown.

The removal of intronic sequences of precursor mRNAs (pre-mRNAs) is catalysed by the spliceosome, a dynamic assembly of five snRNAs and numerous protein factors. The U1, U2, U4, U5 and U6 spliceosomal snRNAs play a fundamental role in pre-mRNA splicing: they select the correct splice sites and orchestrate formation of the active spliceosome (Burge *et al.*, 1999; Will and Lührmann, 2001). The spliceosomal snRNAs contain many post-transcriptionally synthesized 2′-*O*-ribose-methylated nucleotides and pseudouridines (Reddy and Busch, 1988; Massenet *et al.*, 1998). These modified nucleotides are confined to the functionally important regions of snRNAs, which are known to interact with pre-mRNAs, other snRNAs or spliceosomal proteins. Most likely, the modified nucleotides facilitate formation of the specific and dynamic molecular interactions required for the correct function of the spliceosome. Consistent with this view, modification of the U2 snRNA is fundamental for the assembly of the active spliceosome (Pan and Prives, 1989; Ségault *et al.*, 1995; Yu *et al.*, 1998).

The mechanism responsible for the site-specific synthesis of modified nucleotides in snRNAs has long remained an enigma. Recently, it has been established that the nucleolus contains an enormous number of snoRNAs that direct 2′-*O*-methylation and pseudouridylation of rRNAs (Smith and Steitz, 1997; Kiss, 2001). The 2′-*O*-methylation guide snoRNAs carry the conserved 5′-terminal C (RUGAUGA) and 3′-terminal D (CUGA) boxes as well as imperfect internal copies of these elements, called C′ and D′ boxes. To select the correct methylation site, sequences preceding the D or D′ box of the snoRNA form a 10–21 bp helix with rRNAs (Cavaillé *et al.*, 1996; Kiss-László *et al.*, 1996). The pseudouridylation guide snoRNAs fold into a consensus 'hairpin–hinge–hairpin–tail' structure and contain the conserved H (ANANNA) and ACA boxes (Balakin *et al.*, 1996; Ganot *et al.*, 1997b). They possess bipartite rRNA recognition motifs, which form two short helices with ribosomal sequences preceding and following the substrate uridine (Ganot *et al.*, 1997a; Ni *et al.*, 1997). The conserved box C/D and H/ACA motifs play an essential role in snoRNA-guided RNA modification reactions as well as in the processing and nucleolar accumulation of snoRNAs. The C/D and H/ACA boxes accomplish these jobs through binding of specific snoRNP proteins. Thus far, four box C/D (fibrillarin, Nop56p, Nop58p and Snu13p) and four box H/ACA (Garlp, dyskerin, Nhp2p and Nop10p) snoRNA-associated proteins have been identified (Kiss, 2001). Most likely, fibrillarin catalyses the box C/D snoRNA-guided 2′-*O*-methylation reaction (Tollervey *et al.*, 1993; Wang *et al.*, 2000) and dyskerin is the pseudouridine synthase in box H/ACA snoRNPs (Lafontaine *et al.*, 1998; Zebarjadian *et al.*, 1999).

Recent evidence indicates that synthesis of the eight 2′-*O*-methylated nucleotides and the three pseudouridines in the RNA polymerase (pol) III-synthesized U6 spliceosomal snRNA takes place in the nucleolus and that it is directed by box C/D and H/ACA snoRNAs (Tycowski *et al.*, 1998; Ganot *et al.*, 1999). The U6 snRNA, which is transcribed in the nucleoplasm, is transported to the nucleolus to undergo snoRNA-mediated modification (Lange and Gerbi, 2000). Less is known about modification of the pol II-transcribed U1, U2, U4 and U5 snRNAs. Following nucleoplasmic synthesis, these RNAs are transported to the cytoplasm to associate with seven Sm proteins and undergo cap hypermethylation and 3′-end formation (Will and Lührmann, 2001). The newly assembled snRNPs are re-imported to the nucleoplasm, where they accumulate in interchromatin granule clusters (Lamond and Ernshow, 1998; Sleeman and Lamond, 1999b; Matera, 1999; Lewis and Tollervey, 2000). However, a fraction of snRNPs accumulates transiently in Cajal bodies (Carmo-Fonseca *et al.*, 1992; Matera and Ward, 1993; Spector 1993; Sleeman and Lamond, 1999a) and, under certain conditions, in the nucleolus (Lyon *et al.*, 1997; Sleeman and Lamond, 1999a; Lange and Gerbi, 2000; Yu *et al.*, 2001), raising the possibility that these nuclear bodies might have a function in the biogenesis, trafficking or storage of snRNPs.

Microinjection experiments demonstrated that internal modification of the U2 snRNA occurs in the nucleus rather than in the cytoplasm of *Xenopus* oocytes (Yu *et al.*, 2001). The same experiments favoured the nucleolus as the intranuclear locale for U2 modification. This conclusion was seemingly supported by identification of the U85 box C/D–H/ACA composite guide RNA that functions in both 2′-*O*-methylation and pseudouridylation of the U5 snRNA (Jády and Kiss, 2001). In this study, we have characterized six novel putative guide RNAs that are predicted to direct 2′-*O*-methylation and pseudouridylation of the RNA pol II-specific U1, U2, U4 and U5 snRNAs. Unexpectedly, the new guide RNAs as well as the previously characterized U85 RNA accumulate specifically in Cajal bodies and therefore they comprise a novel class of small nuclear RNAs, called the small Cajal body-specific RNAs (scaRNAs). These results indicate that the Cajal body functions in the post-transcriptional modification of pol II-specific spliceosomal snRNAs.

Results

Identification of novel composite box C/D–H/ACA RNAs

We have recently characterized an unusual modification guide RNA, the U85 RNA, which functions both in 2′-*O*-methylation and pseudouridylation of the U5 snRNA (Jády and Kiss, 2001). The U85 RNA is composed of a box C/D and a box H/ACA snoRNA-like domain (Figure 1), and is associated with both fibrillarin and Garlp, which are specific protein components of box C/D and box H/ACA snoRNPs, respectively. To identify novel box C/D–H/ACA composite RNAs, human HeLa RNAs were immunoprecipitated with anti-fibrillarin antibody and size-fractionated on a sequencing gel. RNAs migrating between labelled DNA markers of 234 and 368 nucleotides were recovered and used as a template for cDNA synthesis (Kiss-László *et al.*, 1996). In brief, the 5′ and 3′ termini of the purified RNAs were tagged with a phosphorylated oligonucleotide by using T4 RNA ligase. The double-tagged RNAs were amplified by RT–PCR and cloned into a plasmid vector. Sequence analyses of several individual clones identified three putative snoRNAs, called U87, U88 and U89, which possessed C and D box motifs, but otherwise lacked significant similarity to any known RNAs (Figure 1). Northern blot analyses confirmed that the new RNAs are expressed in HeLa cells (data not shown). The genomic copies of the new RNAs were found within introns of putative protein-coding genes, indicating that they are generated by intron processing (see legend to Figure 1).

A computer-aided structural analysis revealed that the U87, U88 and U89 RNAs can be folded into a two-dimensional structure that is highly reminiscent of the architecture of the U85 box C/D–H/ACA RNA characterized previously (Figure 1). In each RNA, the 5′- and 3′-terminal regions carrying the C and D boxes can form a long stem structure, whereas the middle region folds into the consensus 'hairpin–hinge–hairpin–tail' structure of box H/ACA snoRNAs. Moreover, the single-stranded hinge and tail regions carry potential box H and ACA sequences, respectively. To assess whether the novel putative box C/D–H/ACA snoRNAs are associated with both box C/D and H/ACA snoRNP proteins, immunoprecipitation experiments were performed from a human HeLa cell extract using anti-fibrillarin and anti-GAR1 antibodies (Figure 2A). RNase A/T1 protection analyses demonstrated that the U87, U88 and U89 RNAs, like the U85 box C/D–H/ACA RNA, were precipitated by both antibodies. Please notice that immunological properties of the human U90 and U92 RNPs, although tested in this experiment, will be discussed later. As expected, the U19 box H/ACA snoRNP was precipitated only by the anti-GAR1 and the U3 box C/D snoRNP was recognized only by the anti-fibrillarin antibody and neither of the two antibodies reacted with the U4 spliceosomal snRNP. We concluded that the newly identified U87, U88 and U89 RNAs, together with U85, comprise a new group of snoRNAs that are composed of a box C/D and a H/ACA snoRNA domain, and are associated with both box C/D and box H/ACA snoRNP proteins.

Box C/D–H/ACA RNAs can direct post-transcriptional modification of pol II-specific spliceosomal snRNAs

Examination of the sequences of the new box C/D–H/ACA RNAs showed that they lack significant sequence complementarity to rRNA sequences, making it unlikely that they function in rRNA modification. We have recently demonstrated that the human U85 box C/D–H/ACA RNA directs 2′-*O*-methylation of the C45 and pseudouridylation of the U46 residues in the U5 spliceosomal snRNA (Figure 3). We therefore investigated whether the novel box C/D–H/ACA RNAs could also function in snRNA modification. We found that sequences preceding the D and the putative D′ boxes of the U87 RNA could, in principle, direct 2′-*O*-methylation of the A65 and U41 residues in the U4 and U5 snRNAs, respectively (Figure 3). Indeed, both of these nucleotides are known to be 2′-*O*-methylated in vertebrate U4 and U5 snRNAs (Reddy and

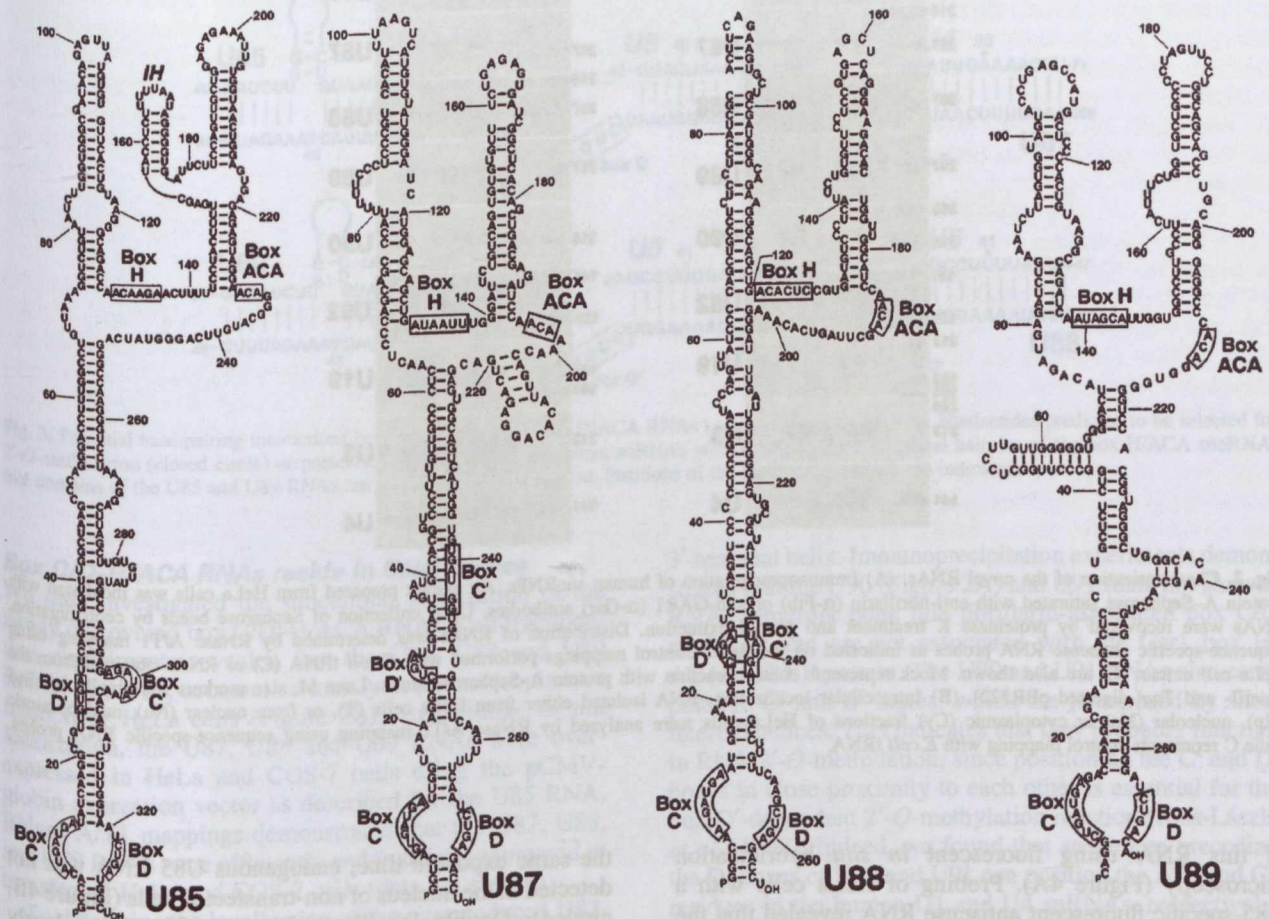


Fig. 1. Computer-predicted two-dimensional structures of human box C/D–H/ACA snoRNAs. The evolutionarily conserved box C, D, H, ACA and the putative box C′ and D′ sequence motifs are boxed. The structure of the U85 snoRNA has been adopted from Jády and Kiss (2001). The genomic copies of the U85 and U89 snoRNAs are in the fourth and sixth introns of the structural maintenance of chromatin (D63880) and the B cell-associated protein (U75511) genes, respectively. The U87 and U88 snoRNAs are encoded within the ninth and twelfth introns of a putative protein-coding gene (BC000061).

Busch, 1988; Massenet *et al.*, 1998). Like U87, the U88 RNA was also predicted to function in 2′-*O*-methylation of the U41 residue in the U5 snRNA, although apart from the putative U5 recognition elements, the two RNAs show no significant sequence similarity. Likewise, the U89 RNA and the previously characterized U85 RNA also seem to share a common function, namely directing pseudouridylation of the U46 residue in the U5 snRNA. Again, the U85 and U89 RNAs, disregarding the short U5 recognition motifs, possess no overall sequence similarity. Our observations strongly suggest that the box C/D–H/ACA RNAs function in the modification of pol II-specific spliceosomal snRNAs and that synthesis of 2′-*O*-methylated nucleotides and pseudouridines in snRNAs is frequently achieved by functionally redundant snoRNPs.

Human U85 RNA accumulates in Cajal bodies

So far, all small RNAs carrying the box C/D or H/ACA motifs have been found to accumulate in the nucleolus (Smith and Steitz, 1997; Tollervey and Kiss, 1997; Weinstein and Steitz, 1999). In fact, the conserved C/D and H/ACA boxes are the key determinants of the nucleolar targeting of snoRNAs (Lange *et al.*, 1998; Samarsky *et al.*, 1998; Narayanan *et al.*, 1999a, b). Therefore, the U85, U87, U88 and U89 box C/D–H/ACA RNAs were assumed to accumulate in the nucleolus. However, during exploration of cDNA libraries of human nucleoplasmic or nucleolar RNAs, we noticed that partial or full-length cDNAs of the U85 RNA frequently appeared in the nucleoplasmic library, but were not found in the nucleolar one (our unpublished results). To test the significance of this unexpected observation, human HeLa cells were fractionated into nuclear, nucleoplasmic, nucleolar and cytoplasmic fractions, and distribution of the U85 RNA was determined by RNase A/T1 mapping (Figure 2B). To our surprise, the U85 RNA was found mostly in the nucleoplasmic fraction. As expected, the U19 box H/ACA and the U3 box C/D snoRNAs showed a predominantly nucleolar accumulation, while the U4 spliceosomal snRNA appeared mostly in the nucleoplasmic fraction.

The fascinating discovery that the U85 box C/D–H/ACA RNA co-purifies with the nucleoplasmic fraction of HeLa cells prompted us to examine further the localization

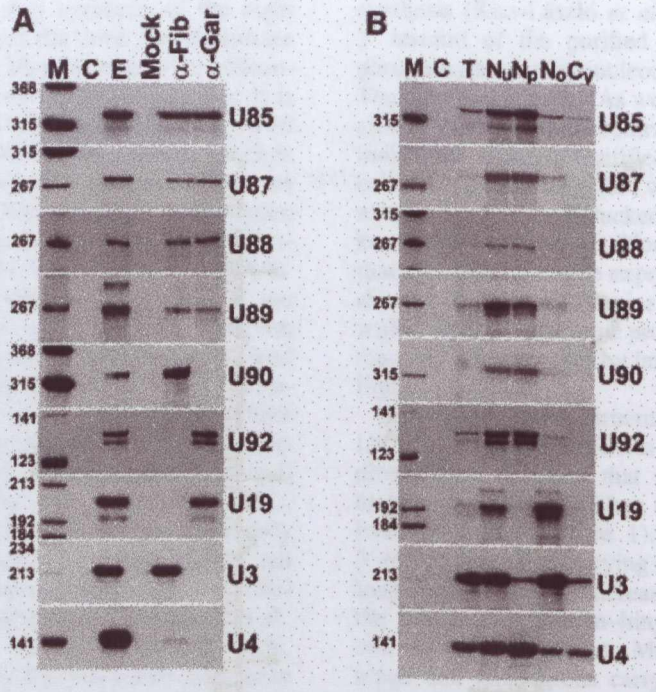


Fig. 2. Characterization of the novel RNAs. (**A**) Immunoprecipitation of human snoRNPs. An extract prepared from HeLa cells was incubated with protein A–Sepharose saturated with anti-fibrillarin (α-Fib) or anti-GAR1 (α-Gar) antibodies. Upon collection of Sepharose beads by centrifugation, RNAs were recovered by proteinase K treatment and phenol extraction. Distribution of RNAs was determined by RNase A/T1 mapping using sequence-specific antisense RNA probes as indicated on the right. Control mappings performed with *E.coli* tRNA (C) or RNAs obtained from the HeLa cell extract (E) are also shown. Mock represents control reaction with protein A–Sepharose alone. Lane M, size markers (terminally labelled *Hae*III- and *Taq*I-digested pBR322). (**B**) Intracellular localization. RNA isolated either from HeLa cells (T), or from nuclear (Nu), nucleoplasmic (Np), nucleolar (No) or cytoplasmic (Cy) fractions of HeLa cells were analysed by RNase A/T1 mapping using sequence-specific RNA probes. Lane C represents control mapping with *E.coli* tRNA.

of this RNA using fluorescent *in situ* hybridization microscopy (Figure 4A). Probing of HeLa cells with a U85-specific fluorescent antisense RNA revealed that the U85 RNA, instead of showing a uniform distribution, localizes to a few sharp, dot-like structures in the nucleoplasm. Since this distribution pattern of U85 was highly reminiscent of that of Cajal bodies (Matera, 1998; Gall, 2000), the cells were also stained with an antibody directed against p80 coilin, a protein marker of the Cajal body (Andrade *et al.*, 1991). The U85 RNA and p80 coilin co-localized perfectly, demonstrating that the nucleoplasmic dots revealed by the U85 probe correspond to Cajal bodies.

To exclude the possibility that a fraction of U85 RNA accumulates outside the Cajal bodies, we investigated the distribution of U85 RNA after overproduction in HeLa cells. To this end, the human U85 RNA gene was inserted into the second intron of the human β-globin gene, which had been placed under the control of the cytomegalovirus (CMV) promoter (Figure 4B). The human U85 RNA was efficiently and correctly expressed in human HeLa and simian COS-7 cells transfected with the pCMV-globin-U85 expression construct (Jády and Kiss, 2001; data not shown). The *in situ* labelling patterns of the U85 RNA probe and the p80 coilin antibody were compared in transfected HeLa cells (Figure 4B). Upon hybridization with the U85 fluorescent probe, a strong signal was observed in cells that overproduced the U85 RNA. With the same exposure time, endogenous U85 RNA was not detected in the nucleus of non-transfected cells (Figure 4B, circled). Double labelling with anti-p80 coilin clearly showed that even after its overproduction, the U85 RNA accumulated exclusively in Cajal bodies. Since RNase A/T1 mappings failed to detect 5′- and/or 3′-extended precursors of U85 (Jády and Kiss, 2001; data not shown), we concluded that the U85-specific sequences accumulating in the Cajal bodies represent fully processed U85 RNAs. As a control, the human mgU6-53 box C/D snoRNA, which directs 2′-*O*-methylation of the RNA pol III-transcribed U6 spliceosomal snRNA (Ganot *et al.*, 1999), was also overproduced in HeLa cells. *In situ* hybridization revealed that the mgU6-53 snoRNA accumulated in large areas of the nucleus, as well as in small dots. It clearly co-localized with a green fluorescent protein (GFP)-tagged version of human fibrillarin, an abundant nucleolar protein that is also present in Cajal bodies (Gall, 2000). Therefore, the mgU6-53 snoRNA accumulated mainly in the nucleolus of transfected cells and, to a smaller extent, in the Cajal bodies, as confirmed by staining with the anti-coilin antibody. This observation was fully consistent with our previous cell fractionation experiments, in which the mgU6-53 snoRNA was localized mainly to the nucleolar fraction of human HeLa cells (Ganot *et al.*, 1999). In summary, we concluded that in both transfected and non-transfected HeLa cells, mature U85 RNA localizes specifically to Cajal bodies.

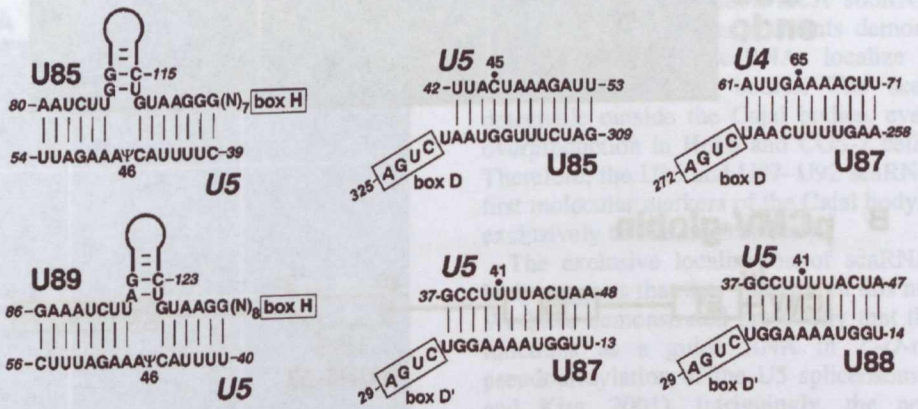


Fig. 3. Potential base-pairing interactions between human box C/D–H/ACA RNAs and spliceosomal snRNAs. Nucleotides predicted to be selected for 2′-*O*-methylation (closed circle) or pseudouridylation (Ψ) in the human snRNAs are indicated. The 5′-terminal hairpins of the box H/ACA snoRNA-like domains of the U85 and U89 RNAs are schematically represented. Positions of the H, D and D′ boxes are indicated.

Box C/D–H/ACA RNAs reside in Cajal bodies

We then investigated the subcellular localization of the newly discovered box C/D–H/ACA RNAs. Cell fractionation experiments revealed that the U87, U88 and U89 RNAs, like U85, accumulate mostly in the nucleoplasmic fraction of HeLa cells (Figure 2B). To facilitate *in situ* localization, the U87, U88 and U89 RNAs were overexpressed in HeLa and COS-7 cells using the pCMV-globin expression vector as described for the U85 RNA. RNase A/T1 mappings demonstrated that the U87, U88, and U89 RNAs were efficiently and faithfully expressed in transfected HeLa and COS-7 cells (data not shown).

To determine the localization of the overproduced U87, U88 and U89 RNAs in HeLa cells, double labelling experiments were carried out with sequence-specific fluorescent RNA (U87, U89) or oligonucleotide (U88) probes and the anti-p80 coilin antibody (Figure 5). For each RNA, a perfect co-localization was observed with coilin, demonstrating that the overexpressed RNAs accumulate in the Cajal bodies of HeLa cells. The same results were obtained by *in situ* localization of the transiently expressed human U87, U88 and U89 RNAs in COS-7 cells (data not shown). Moreover, probing of non-transfected HeLa cells with the U88 probe specifically stained Cajal bodies, indicating that endogenous U88 RNA also resides in this nuclear organelle. Our results demonstrate that the human U85, U87, U88 and U89 box C/D–H/ACA RNAs comprise a novel class of cellular RNAs that localize specifically to the Cajal body. The new group of snRNAs was collectively named small Cajal body-specific RNAs (scaRNAs).

Guide RNAs implicated in modification of pol II-specific spliceosomal snRNAs reside in Cajal bodies

The results presented thus far indicate that scaRNAs directing modification of snRNAs are composed of a box C/D and a H/ACA snoRNA domain. In cDNA libraries of human fibrillarin- and Gar1p-associated snoRNAs, we have identified three novel RNAs, called U90, U91 and U92 (Figure 6A). The U90 and U91 RNAs possess the characteristic core motif of box C/D snoRNAs, which includes the conserved C and D boxes and a short 5′, 3′-terminal helix. Immunoprecipitation experiments demonstrated that the U90 (Figure 2A) and U91 (data not shown) RNAs represent bona fide fibrillarin-associated box C/D snoRNAs and are not associated with the Gar1 box H/ACA snoRNP protein. The U90 and U91 RNAs also carry putative C′ and D′ boxes, which are juxtaposed by short internal helices. This indicates that they probably function in RNA 2′-*O*-methylation, since positioning the C′ and D′ boxes in close proximity to each other is essential for the box D′-dependent 2′-*O*-methylation reaction (Kiss-László *et al.*, 1998). Indeed, we found that sequences preceding the D′ boxes of U90 and U91 can position the A70 and C8 residues in the human U1 and U4 snRNAs, respectively, for 2′-*O*-methylation (Figure 6A). Both of these nucleotides are in fact 2′-*O*-methylated in mammalian U1 and U4 snRNAs (Reddy and Busch, 1988). In contrast to U90 and U91, the human U92 RNA lacks C and D boxes, but it folds into the consensus 'hairpin–hinge–hairpin–tail' structure of box H/ACA snoRNAs (Figure 6A). The U92 RNA binds the Gar1, but not the fibrillarin snoRNP protein (Figure 3A), demonstrating that it is an authentic box H/ACA snoRNA. We noticed that the U92 box H/ACA RNA, in theory, is capable of directing pseurouridylation of the U44 residue in the U2 snRNA (Figure 6A). During the course of this study, a large-scale characterization of mouse non-coding nuclear RNAs revealed partial sequences of the apparent mouse orthologues of the human U91 (MBII-119) and U92 (MBI-57) RNAs, showing that these putative snRNA modification guide RNAs are conserved in mammalian cells (Hüttenhofer *et al.*, 2001).

To assess whether all guide RNAs implicated in modification of pol II-specific snRNAs accumulate in Cajal bodies, we investigated the intracellular localization of the newly identified U90, U91 and U92 RNAs. Encouragingly, cell fractionation experiments revealed a nucleoplasmic localization for the U90 box C/D and the U92 box H/ACA RNAs (Figure 2B). Next, probing of HeLa cells with fluorescent antisense probes specific for the U91 and U92 RNAs resulted in strongly staining dots in the nucleoplasm (Figure 6B, U91endo and U92endo). Double labelling with anti-p80 coilin antibody showed that the dots highlighted by the U91- and U92-specific probes corresponded to Cajal bodies. The U92 RNA,

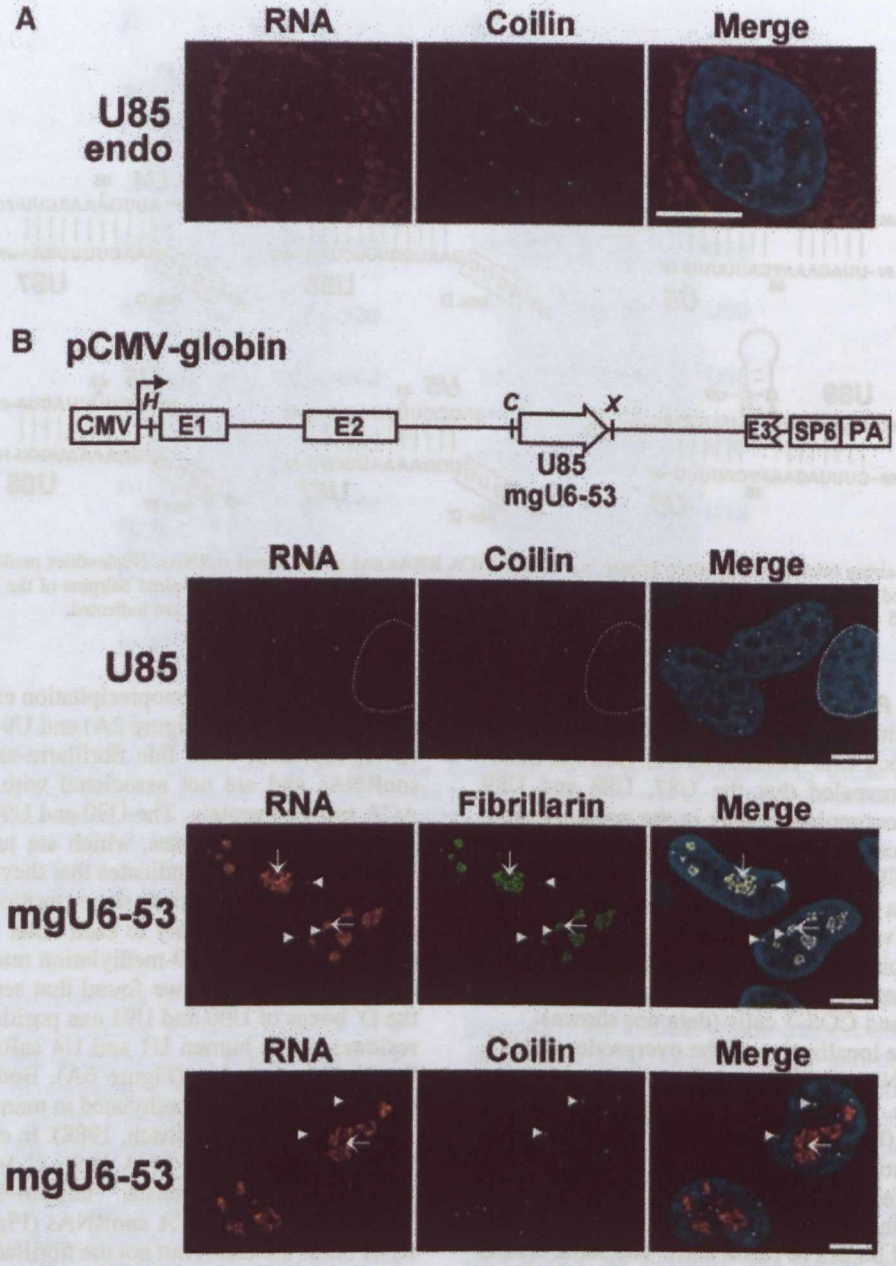


Fig. 4. Human U85 box C/D–H/ACA RNA localizes to Cajal bodies. (**A**) *In situ* localization of the endogenous U85 RNA. Human HeLa cells were probed with a fluorescent antisense RNA complementary to the human U85 RNA from position 1 to 76 and with an antibody directed against p80 coilin as indicated above the panels. The merged image shows that the U85 RNA co-localizes with p80 coilin. Scale bar, 10 μm. (**B**) *In situ* localization of transiently overexpressed U85 and mgU6-53 RNAs in HeLa cells. The schematic structure of the pCMV-globin expression construct is shown. The exons (E1, E2 and E3) and the polyadenylation site (PA) of the human β-globin gene are indicated. The transcription initiation site of the cytomegalovirus promoter (CMV) is indicated. The coding regions of the human U85 and mgU6-53 RNAs in the second intron of the globin gene are represented by an open arrow. Relevant restriction sites are shown (H, *Hin*dIII; C, *Cla*I; X, *Xho*I). The pCMV-globin-mgU6-53 plasmid was co-transfected with an expression construct producing a GFP-tagged version of the human fibrillarin protein. The overexpressed U85 and mgU6-53 RNAs were visualized by sequence-specific fluorescent RNA probes and Cajal bodies were specifically stained by anti-p80 coilin. Arrows indicate nucleolar stained structures enriched in mgU6-53 and fibrillarin. Arrowheads point to Cajal bodies, which clearly contain mgU6-53, fibrillarin and p80 coilin. The contour of the nucleus of a non-transfected HeLa cell is highlighted.

probably due to its low abundance, was not directly detectable in HeLa cells. Therefore, localization of this RNA and also the U90 RNA was investigated in HeLa cells transfected with pCMV-globin constructs expressing the U92 and U90 RNAs. U90 and U92 were correctly and efficiently expressed in transfected HeLa cells (data not shown) and, as *in situ* localization experiments demonstrated, localized specifically to Cajal bodies (Figure 6B, U90 and U92). We concluded that 2′-*O*-methylation and pseudouridylation guide RNAs that function in post-transcriptional modification of the RNA pol II-transcribed spliceosomal snRNAs, irrespective of their box elements, associated proteins and molecular sizes, specifically and exclusively accumulate in Cajal bodies.

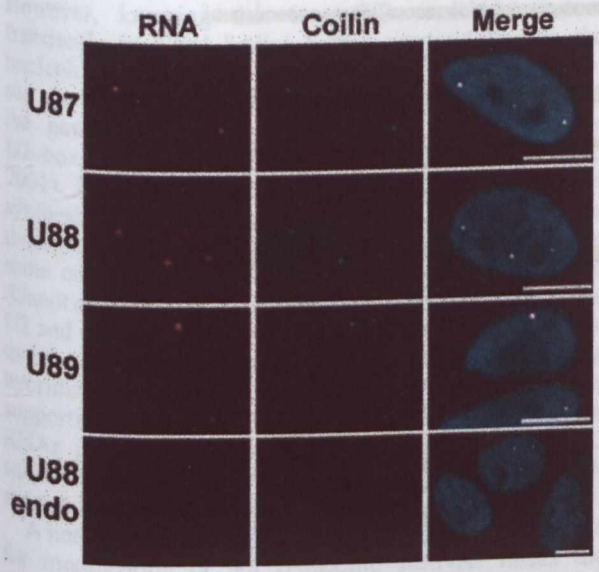


Fig. 5. *In situ* localization of human U87, U88 and U89 box C/D–H/ACA RNAs. Human U87, U88 and U89 RNAs were transiently over expressed in HeLa cells using the pCMV-globin expression construct. The intracellular distribution of the overproduced U87, U88 and U89 RNAs as well as the endogenous HeLa U88 RNA (U88endo) was investigated by fluorescent *in situ* hybridization. Cajal bodies were visualized by staining with anti-p80 coilin antibody. For other details, see the legend to Figure 4.

Discussion

The interphase nucleus has an intricate structural and functional organization. It contains many distinct subdomains, also known as nuclear bodies, which represent specific and dynamic assemblies of protein and RNA factors involved in various aspects of nuclear gene expression. The Cajal body is a prominent and extensively studied nuclear organelle of unknown function (Bohmann *et al.*, 1995; Lamond and Earnshaw, 1998; Matera, 1999; Gall, 2000; Lewis and Tollervey, 2000). Understanding of the function of the Cajal body has been hampered largely by the fact that all its known components are also distributed in the nucleoplasm or, frequently, they concentrate in other nuclear bodies. Thus far, the p80 coilin autoantigene is the only unambiguous molecular marker for the Cajal body, although most of this protein is dispersed in the nucleoplasm. The Cajal body is enriched in spliceosomal snRNPs and nucleolar snoRNPs, but lacks mRNAs and rRNAs, indicating that it does not function in mRNA splicing or rRNA maturation (Bohmann *et al.*, 1995). Several lines of evidence indicate that there is a flux of newly synthesized snRNPs and snoRNPs through the Cajal body, leading to the idea that the Cajal body may function in the biogenesis and/or transport of snRNPs and snoRNPs (Sleeman and Lamond, 1999a; Gall, 2000).

In this report, we have described a novel class of human small nuclear RNAs, scaRNAs, which accumulate specifically in Cajal bodies. Apart from the fact that scaRNAs frequently contain both box C/D and H/ACA snoRNA domains, they are structurally indistinguishable from the canonical box C/D and box H/ACA snoRNAs (Figures 1 and 6A). Depending on their box elements, scaRNAs are associated with characteristic protein components of the nucleolar box C/D and H/ACA snoRNPs (Figure 2A). *In situ* hybridization experiments demonstrated that the U85 and U87–U92 scaRNAs localize specifically and exclusively to Cajal bodies. The scaRNAs are not detectable outside the Cajal bodies, even after massive overproduction in HeLa and COS-7 cells (Figures 4–6). Therefore, the U85 and U87–U92 scaRNAs represent the first molecular markers of the Cajal body, which partition exclusively to this nuclear body.

The exclusive localization of scaRNAs to the Cajal body suggests that they function in this nuclear organelle. We have demonstrated previously that the U85 scaRNA functions as a guide RNA in 2′-*O*-methylation and pseudouridylation of the U5 spliceosomal snRNA (Jády and Kiss, 2001). Intriguingly, the newly discovered scaRNAs are also predicted to direct 2′-*O*-methylation and/or pseudouridylation of the U1, U2, U4 and U5 snRNAs (Figures 3 and 6A). We propose that post-transcriptional modification of pol II-specific spliceosomal snRNAs is directed by a novel class of guide RNAs that reside in the Cajal body. Mammalian U1, U2, U4 and U5 snRNAs together carry 13 2′-*O*-methyl groups and 21 pseudouridine residues (Reddy and Busch, 1988). In this study, seven scaRNAs have been connected with the synthesis of five 2′-*O*-methylated nucleotides (U85, U87, U88, U90 and U91) and two pseudouridine residues (U85, U89 and U92) in pol II-specific snRNAs. Currently, we have identified another box H/ACA scaRNA, which is predicted to direct pseudouridylation of the U2 snRNA at the U54 position (our unpublished data). In a previous study, two box C/D (MBII-19 and MBII-382) and two box H/ACA (MBI-57 and MBI-125) RNAs have been implicated in the synthesis of three 2′-*O*-methylated nucleotides and two pseudouridines in the U2 snRNA (Hüttenhofer *et al.*, 2001). Unfortunately, the cellular localization of these RNAs has not been explored. Nevertheless, thus far, 12 putative guide scaRNAs have been linked with the synthesis of 12 2′-*O*-methylated nucleotides and two pseudouridines in the U1, U2, U4 and U5 snRNAs. We can envisage that the synthesis of most, if not all, 2′-*O*-methylated nucleotides and pseudouridines in pol II-specific spliceosomal snRNAs is directed by scaRNAs.

So far, the Cajal body is the only common nuclear locale where both spliceosomal snRNAs and their modification guide RNAs have been demonstrated to accumulate. Therefore, it seems unlikely that the Cajal body is only a storage place for scaRNAs, and scaRNA-directed snRNA modification occurs in another nuclear compartment. Previously, the nucleolus has been implicated in post-transcriptional modification of both the RNA pol III-transcribed U6 (Tycowski *et al.*, 1998; Ganot *et al.*, 1999; Lange and Gerbi, 2000) and the pol II-specific U2 spliceosomal snRNAs (Yu *et al.*, 2001). The idea that modification of the U6 snRNA occurs within the nucleolus is supported by the fact that all *trans*-acting factors, most likely snoRNPs, that accomplish the synthesis of the eight 2′-*O*-methylated nucleotides and the three pseudouridines of the U6 snRNA are present and are functionally active in the nucleolus (Tycowski *et al.*, 1998; Ganot *et al.*, 1999; Figure 4B). Upon microinjection into the nucleoplasm of *Xenopus* oocyte, the U6 snRNA appears transiently in the nucleolus, suggesting that it transits through the nucleolus to undergo snoRNA-directed modification (Lange and

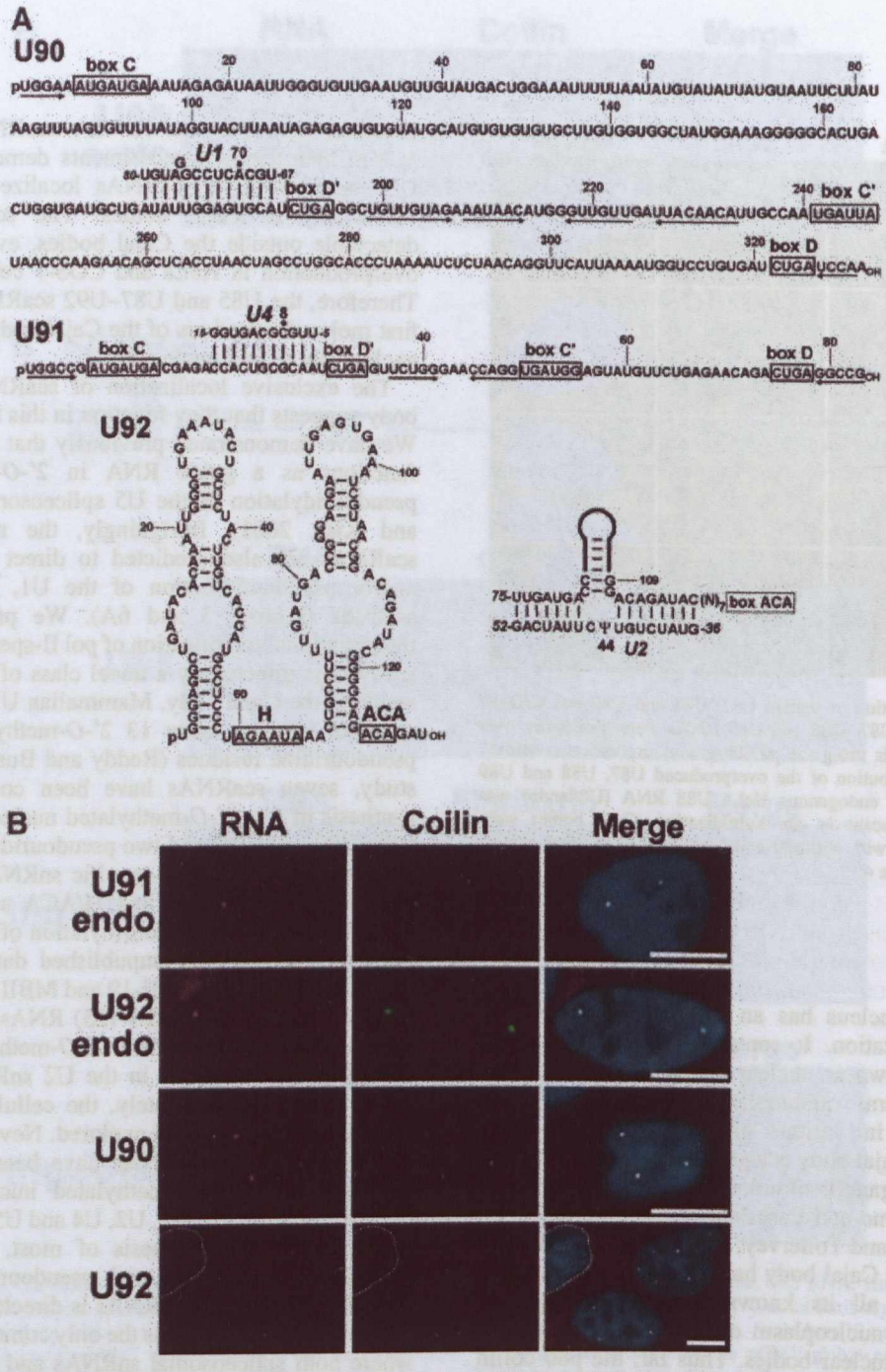


Fig. 6. Localization of box C/D and H/ACA RNAs predicted to function in modification of pol II-specific spliceosomal snRNAs. (**A**) Structure and function of human U90, U91 box C/D and U92 box H/ACA RNAs. The conserved box C/D and the potential C′ and D′ motifs of the U90 and U91 RNAs are indicated. Inverted arrows under the sequences of U90 and U91 indicate nucleotides predicted to form terminal and internal helices. The predicted two-dimensional structure of the U92 RNA was obtained by computer folding. The H and ACA sequence motifs are boxed. Predicted base-pairing interactions of the U90, U91 and U92 RNAs with the U1, U4 and U2 snRNAs, respectively, are indicated. The 2′-*O*-methylated A70 and C8 nucleotides in the selected sequences of the U1 and U4 snRNAs are indicated by dots. The pseudouridine residue (Ψ) at position 44 in the human U2 snRNA is marked. (**B**) Localization of human U90, U91 and U92 RNAs in HeLa cells. Overexpressed (U90 and U92) and endogenous (U91endo and U92endo) modification guide RNAs were visualized with sequence-specific fluorescent probes. For other details, see legends to Figures 4 and 5.

Gerbi, 2000). Although maturation of the RNA pol II-synthesized U1, U2, U4 and U5 snRNAs includes a cytoplasmic phase (Will and Lührmann, 2001), internal modification of the pol II-transcribed U2 snRNA occurs in the nucleus (Yu *et al.*, 2001). In *Xenopus* oocytes, modification of the U2 snRNA was found to require an intact Sm motif (Yu *et al.*, 2001). Lack of a functional Sm site could be complemented by addition of a 5′,3′-terminal box C/D snoRNA core structure. As the resulting chimeric U2–box C/D RNA was efficiently modified and it accumulated in the nucleolus, it was suggested that the nucleolus provides the cellular locale for U2 modification.

However, knowing that box C/D snoRNAs appear transiently in Cajal bodies before accumulating in the nucleolus (Samarsky *et al.*, 1998; Narayanan *et al.*, 1999b; see also Figure 4B), these experiments could not rule out the possibility that the modification of the chimeric U2–box C/D RNA happened in Cajal bodies (Yu *et al.*, 2001). Previously, we found that U2 snRNA sequences synthesized within the nucleolus by RNA pol I were not modified, but U6-specific sequences expressed under the same conditions were correctly and efficiently modified (Ganot *et al.*, 1999). This indicated that modification of the U2 and probably other pol II-specific snRNAs, in contrast to U6, occurs outside the nucleolus. Consistently, *in situ* hybridization experiments failed to provide any evidence supporting accumulation of the U85 and U87–U92 guide RNAs in the nucleolus, while the mgU6-53 box C/D snoRNA directing 2′-*O*-methylation of the U6 snRNA accumulated mostly in the nucleolus (Figures 4–6).

A notion that the Cajal body provides the cellular locale for modification of pol II-specific snRNAs raises the question of whether the scaRNA-guided modification of snRNAs occurs before or after the cytoplasmic assembly of Sm snRNPs. Upon microinjection into the cytoplasm of *Xenopus* oocytes, *in vitro* synthesized U2 RNA accumulates in the nucleoplasm and undergoes pseudouridylation and 2′-*O*-methylation (Yu *et al.*, 2001). This demonstrates that modification of the U2 snRNA can occur in the nascent Sm snRNP after its re-entry into the nucleus. Indeed, the newly made snRNPs accumulate transiently in Cajal bodies after re-entering the nucleus (Carvalho *et al.*, 1999; Sleeman and Lamond, 1999a). However, we cannot exclude that at least some modifications are introduced into the U2 snRNA before its export to the cytoplasm, since nascent precursor U2 snRNAs have also been reported to appear in Cajal bodies (Smith and Lawrence, 2000).

Since scaRNAs possess all those elements necessary and sufficient to direct the nucleolar accumulation of box C/D and H/ACA snoRNAs (reviewed in Tollervey and Kiss, 1997; Weinstein and Steitz, 1999), the molecular mechanism directing scaRNAs to the Cajal body, rather than to the nucleolus, is conjectural. The U85, U87, U88 and U89 scaRNAs are composed of a box C/D and a box H/ACA snoRNA domain. Therefore, these RNAs, in theory, possess two independent nucleolar localization signals, any of which should be able to direct these RNAs into the nucleolus. Apparently this is not the case. Since box C/D snoRNAs are known to transit through Cajal bodies before accumulating in the nucleolus (Samarsky *et al.*, 1998; Narayanan *et al.*, 1999b), we can hypothesize that the C/D motif targets the box C/D (U90, U91) and the composite box C/D–H/ACA (U85, U87, U88 and U89) scaRNAs to the Cajal body, where a putative retention factor inhibits the nucleolar export of these RNAs. However, identification of the U92 box H/ACA scaRNA clearly demonstrates that RNAs lacking C and D boxes can also accumulate in the Cajal body (Figure 6B). Whether the Cajal body-specific accumulation of the box C/D and box H/ACA scaRNAs is supported by two different mechanisms remains uncertain.

In summary, demonstration that the molecular machinery mediating the 2′-*O*-methylation and pseudouridylation of the RNA pol II-transcribed U1, U2, U4 and U5 spliceosomal snRNAs is sequestered into the Cajal body sheds new light on the cellular function of this mysterious nucleoplasmic organelle and reveals new details of the spatial organization of the biogenesis of spliceosomal snRNPs. In the future, dissection of the *cis*-acting elements and identification of the *trans*-acting factors responsible for the Cajal body-specific accumulation of scaRNAs will provide us with new insights into the molecular mechanism underlying the nuclear compartmentalization of eukaryotic RNA biogenesis.

Materials and methods

General procedures

Unless indicated otherwise, all techniques used for manipulating RNA, DNA and oligonucleotides were according to Sambrook *et al.* (1989). The following oligonucleotides were used in this study: 1, AATAAAGCGG-CCGCGGATCCAA; 2, TTGGATCCGCGGCCGCTTTAT; 3, ATAAT-CGATGGAAGGTGTTTGTTATC; 4, ATACTCGAGTTTCACTCAC-TTCTTTC; 5, ATAATCGATAGTCCCACTCCACTCCTGTG; 6, ATA-CTCGAGGGTGACCAAACCTTTTACCC; 7, ATAATCGATACATC-AGTGAATACCTTCTG; 8, ATACTCGAGAACATCAGGACTCCTT-ATGT; 9, ATAATCGATCTCAGCCCAGCCCCTAGGGC; 10, ATAC-TCGAGCCTGGCCCTGTCCTTACCAC; 11, ATAATCGATTCTCCA-TAACAAGCATTAAT; 12, ATACTCGAGTAACTAATAAGTTTT-ACTCT; 13, ATAATCGATTGGGAGGCTGATACACAAATTGG; 14, ATACTCGAGATCTGTCTGCCCCGTATCTG; 15, ATAATCGATAT-CTCCCAATGGTACCTGAAC; 16, ATACTCGAGTCAGTCATGAT-GGAATGGGG; 17, GCCACATGATGATATCAAGGC; 18, GTAAT-ACGACTCACTATAGGGGACCTTTAACAGGCCAAAGG; 19, TAA-GTCATGTGTATGGGATC; 20, AATACGACTCACTATAGGGGGT-TTGTTGGATACTCGTC; 21, AGATCTGAAATCTTAGTGGT; 22, A-ATACGACTCACTATAGGGGGCAGCACCAGAAATGAAGGC; 23, GCTGATATTGGAGTGCATCTG; 24, AATACGACTCACTATAGG-GGGGTGCCAGGCTAGTTAGGTG; 25, TGGGAGGCTGATACACA-AATTGG; 26, AATACGACTCACTATAGGGGGGATCTGTCTGCC-CCGTATCTG; 27, TCCCAATGATGAGTTGCC; 28, AATACGACT-CACTATAGGGGGACCCCTCAGATCTTCATGTG; 29, CT*GGGA-TGCCGGGAGGGGAT*CTGAGGACT*CAGACCTTTTACCATT*C; 30, CGGCCT*CAGTCAGTTGT*CAGAAGATACT*CCAT*CACC-TGGTTC. Sequences corresponding to the T7 RNA polymerase promoter are underlined. Amino-allyl-modified T residues are marked by asterisks.

Plasmid construction

Synthesis and cloning of cDNAs of human scaRNAs were performed as described by Kiss-László *et al.* (1996). RNAs immunoprecipitated from a HeLa cell extract by anti-fibrillarin or anti-GAR1 antibodies were size-fractionated on a 6% sequencing gel. The appropriate RNA fractions were recovered and incubated with an excess of 5′-end-phosphorylated oligonucleotide 1 in the presence of T4 RNA ligase. The ligation product was used as a template for cDNA synthesis by using oligonucleotide 2 as a primer and AMV reverse transcriptase. The resulting first strand cDNA was used as a template for PCR amplification by Vent polymerase using oligonucleotides 1 and 2 as primers. The amplified DNA was digested by *Bam*HI and inserted into the same site in pBluescribe (Stratagene).

To obtain pCMV-globin, the *Hin*dIII–*Eco*RI fragment of the human β-globin gene carrying three artificial restriction sites (*Cla*I, *Mlu*I and *Xho*I) in its second intron was excised from the pG_{CXM} expression vector (Kiss and Filipowicz, 1995) and inserted into the same sites of the pcDNA3 vector (Invitrogen). Removal of the *Eco*RI–*Xba*I fragment of the resultant pcDNA3-globin construct yielded pCMV-globin. DNA fragments containing the coding genes of the human U85, U87, U88, U89, U90 and U92 scaRNAs were PCR amplified using human genomic DNA as a template, and oligonucleotides 3/4, 5/6, 7/8, 9/10, 11/12 and 13/14 as primers, respectively. The amplified DNAs were digested with *Cla*I and *Xho*I and were inserted into the same sites of pCMV-globin. Likewise, a fragment of the human genome encoding the mgU6-53 snoRNA was amplified (oligonucleotides 15 and 16), and cloned into the *Cla*I–*Xho*I sites of pCMV-globin. Transfection of human HeLa and COS-7 cells was performed as described before, except that the transfection reagent Fugene (Roche) was used (Kiss and Filipowicz, 1995).

RNA extraction and analysis

Guanidinium thiocyanate/phenol–chloroform extraction was used to isolate RNA from human HeLa and simian COS-7 cells, and from the nuclear, nucleolar and nucleoplasmic fractions of HeLa cells (Goodall *et al.*, 1990). RNAs were extracted from the cytoplasmic fraction of HeLa cells and the Sepharose beads of immunoprecipitation reactions by proteinase K treatment followed by phenol–chloroform extraction. RNase A/T1 protection assays were performed as described (Goodall *et al.*, 1990). Synthesis of antisense RNA probes for mapping of U4, U3 and U19 RNAs have been described (Ganot *et al.*, 1997b). For synthesis of sequence-specific probes for the U85, U87, U88, U89, U90, U91 and U92 scaRNAs, the appropriate pCMV-globin expression construct was linearized by *Hin*dIII and used as template for transcription by the SP6 RNA polymerase. After synthesis, each RNA probe was purified on a 6% sequencing gel.

Immunoprecipitation and cell fractionation

Human HeLa S3 cells were grown in a suspension culture in Joklik's modified Eagle's medium (Gibco) containing 5% newborn calf serum. About 5×10^6 cells were washed in ice-cold TBS (150 mM NaCl, 40 mM Tris–HCl pH 7.4), sonicated in 200 mM NaCl, 40 mM Tris–HCl pH 7.4 containing 0.05% Nonidet P-40 as reported by Tyc and Steitz (1989) and centrifuged for 10 min at 10 000 *g* at 4°C. The supernatant was incubated with 2 mg of protein A–Sepharose beads swollen in the sonication buffer and saturated with anti-fibrillarin (72B9) or anti-hGAR1 antibodies, which were kindly provided by Drs J.A.Steitz and W.Filipowicz, respectively. After immunoprecipitation, the beads were washed four times with 10 vol of sonication buffer. Isolation of nuclei from HeLa cells, and fractionation of nuclei into nucleoplasmic and nucleolar fractions were performed as described (Tyc and Steitz, 1989).

Preparation of probes for in situ hybridization

Synthesis and labelling of antisense RNA probes were performed according to the protocol of Dr R.Singer (http://singerlab.aecom.yu.edu). To produce probes specific for U85, U87, U89, U90, U92 and mgU6-53 RNAs, fragments of these RNAs were amplified by PCR with oligonucleotides 17/18, 19/20, 21/22, 23/24, 25/26 and 27/28, respectively. The 3′-end-specific primers carried the sequence of the T7 RNA polymerase promoter. The resulting DNA fragments were used as templates for *in vitro* transcription by T7 RNA polymerase in the presence of 5-(3-aminoallyl) uridine 5′-triphosphate. About 2–5 µg of RNA purified on a 6% sequencing gel was resuspended in 70 µl of 0.1 M Na-carbonate buffer pH 8.8 and mixed with 30 µl of dimethylsulfoxide containing a vial of FluoroLink Cy3-monofunctional dye (Amersham). Labelling was performed for 48 h in the dark, at room temperature, with occasional vortexing. The unreacted dye was removed by ethanol precipitation of the RNA. Specific activity of the probes was determined by absorption spectroscopy. To detect U88 and U91, amino-modified antisense oligonucleotides 29 and 30 were synthesized, respectively, and reacted with Cy3 monofunctional dye.

Fluorescent in situ hybridization and image acquisition and processing

HeLa cells were fixed in phosphate-buffered saline (PBS) buffer (100 mM Na_2HPO_4, 20 mM KH_2PO_4, 137 mM NaCl, 27 mM KCl pH 7.4) containing 4% formaldehyde for 30 min at room temperature. The cells were rinsed twice with PBS and permeabilized by incubation overnight in 70% ethanol. After rehydration in 2 × SSC (300 mM NaCl, 30 mM sodium citrate pH 7.0) containing 50% formamide, cells were hybridized overnight at 37°C in 40 µl of a mixture containing 10% dextran sulfate, 2 mM vanadyl–ribonucleoside complex, 0.02% RNase-free bovine serum albumin (BSA), 40 µg of *Escherichia coli* tRNA, 2 × SSC, 50% formamide, 20 ng of labelled probe. Cells probed with fluorescent antisense RNA were washed twice for 30 min in 0.1 × SSC, 50% formamide at 50°C, while oligonucleotide-treated cells were rinsed with 2 × SSC, 50% formamide, at 37°C. p80 coilin was detected with polyclonal rabbit anti-coilin antibody (1/100 dilution, kindly provided by Dr A.Lamond) followed by incubation with anti-rabbit antibodies conjugated to fluorescein (1/300 dilution, Sigma). Slides were incubated for 1 h at 37°C in PBS containing 1% BSA and washed twice for 15 min in PBS at room temperature. Slides were mounted in mounting media containing 90% glycerol, 1 × PBS, 0.1 µg/ml of 4′,6-diamidino-2-phenylindole and 1 mg/ml *p*-phenylendiamine. Images were acquired on a DMRA microscope equipped for epifluorescence (Leica), and with a CoolSnap camera (Photometrics) controlled by the software Metamorph (Universal Imaging). Images were then pseudo-coloured with Photoshop (Adobe Systems).

Accession numbers

The DDBJ/EMBL/GenBank accession numbers of human U87, U88, U89, U90, U91 and U93 scaRNAs are AY077737, AY77738, AY77739, AY77740, AY77741 and AY77742, respectively.

Acknowledgements

We thank M.Weber for critical reading of the manuscript. We are grateful to Y.de Preval for synthesis of oligodeoxynucleotides. X.D. and C.V. were supported by la Fondation pour la Recherche Médicale and Association pour la Recherche contre le Cancer, respectively. B.E.J. and A.M.K. were funded by the French Government and the Hungarian Academy of Sciences. This work was supported by grants from Association pour la Recherche contre le Cancer, la Ligue Nationale contre le Cancer, the French MNRT (ACI) and the Hungarian Research Foundation (OTKA, T29042 and T31738).

References

Andrade,L.E., Chan,E.K., Raska,I., Peebles,C.L., Roos,G. and Tan,E.M. (1991) Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. *J. Exp. Med.*, **173**, 1407–1419.

Balakin,A.G., Smith,L. and Fournier,M.J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, **86**, 823–834.

Bohmann,K., Ferreira,J., Santama,N., Weis,K. and Lamond,A.I. (1995) Molecular analysis of the coiled body. *J. Cell Sci.*, **19**, *Suppl.*, 107–113.

Burge,C.B., Tuschl,T. and Sharp,P.A. (1999) Splicing of precursors to mRNAs by the spliceosome. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 525–560.

Carmo-Fonseca,M., Pepperkok,R., Carvalho,M.T. and Lamond,A.I. (1992) Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *J. Cell Biol.*, **117**, 1–14.

Carvalho,T., Almeida,F., Calapez,A., Lafarga,M., Berciano,M.T. and Carmo-Fonseca,M. (1999) The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. *J. Cell Biol.*, **147**, 715–728.

Cavaillé,J., Nicoloso,M. and Bachellerie,J.P. (1996) Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature*, **383**, 732–735.

Gall,J.G. (2000) Cajal bodies: the first 100 years. *Annu. Rev. Cell Dev. Biol.*, **16**, 273–300.

Ganot,P., Bortolin,M.-L. and Kiss,T. (1997a) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.

Ganot,P., Caizergues-Ferrer,M. and Kiss,T. (1997b) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**, 941–956.

Ganot,P., Jády,B.E., Bortolin,M.-L., Darzacq,X. and Kiss,T. (1999) Nucleolar factors direct the 2′-*O*-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol. Cell. Biol.*, **19**, 6906–6917.

Goodall,G.J., Wiebauer,K. and Filipowicz,W. (1990) Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol.*, **181**, 148–161.

Hüttenhofer,A., Kiefmann,M., Meier-Ewert,S., O'Brien,J., Lehrach,H., Bachellerie,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.

Jády,B.E. and Kiss,T. (2001) A small nucleolar guide RNA functions both in 2′-*O*-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J.*, **20**, 541–551.

Kiss,T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**, 3617–3622.

Kiss,T. and Filipowicz,W. (1995) Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev.*, **9**, 1411–1424.

Kiss-László,Z., Henry,Y., Bachellerie,J.-P., Caizergues-Ferrer,M. and Kiss,T. (1996) Site-specific ribose methylation of preribosomal RNA: A novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.

Kiss-László,Z., Henry,Y. and Kiss,T. (1998) Sequence and structural

elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.*, **17**, 797–807.
Lafontaine,D.L., Bousquet-Antonelli,C., Henry,Y., Caizergues-Ferrer,M. and Tollervey,D. (1998) The box H+ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.*, **12**, 527–537.
Lamond,A.I. and Earnshaw,W.C. (1998) Structure and function in the nucleus. *Science*, **280**, 547–553.
Lange,T.S. and Gerbi,S.A. (2000) Transient nucleolar localization of U6 small nuclear RNA in *Xenopus laevis* oocytes. *Mol. Biol. Cell*, **11**, 2419–2428.
Lange,T.S., Borovjagin,A., Maxwell,E.S. and Gerbi,S.A. (1998) Conserved boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. *EMBO J.*, **17**, 3176–3187.
Lewis,J.D. and Tollervey,D. (2000) Like attracts like: getting RNA processing together in the nucleus. *Science*, **288**, 1385–1389.
Lyon,C.E., Bohmann,K., Sleeman,J. and Lamond,A.I. (1997) Inhibition of protein dephosphorylation results in the accumulation of splicing snRNPs and coiled bodies within the nucleolus. *Exp. Cell Res.*, **230**, 84–93.
Massenet,S., Mougin,A. and Branlant,C. (1998) Posttranscriptional modifications in the U small nuclear RNAs. In Grosjean,H. and Benne,R. (eds), *Modification and Editing of RNA*. ASM Press, Washington DC, pp. 201–227.
Matera,A.G. (1998) Of coiled bodies, gems, and salmon. *J. Cell. Biochem.*, **70**, 181–192.
Matera,A.G. (1999) Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell Biol.*, **9**, 302–309.
Matera,A.G. and Ward,D.C. (1993) Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. *J. Cell Biol.*, **121**, 715–727.
Narayanan,A., Lukowiak,A., Jády,B.E., Dragon,F., Kiss,T., Terns,R.M. and Terns,M.P. (1999a) Nucleolar localization signals of box H/ACA small nucleolar RNAs. *EMBO J.*, **18**, 5120–5130.
Narayanan,A., Speckmann,W., Terns,R. and Terns,M.P. (1999b) Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol. Biol. Cell*, **10**, 2131–2147.
Ni,J., Tien,A.L. and Fournier,M.J. (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, **89**, 565–573.
Pan,Z.Q. and Prives,C. (1989) U2 snRNA sequences that bind U2-specific proteins are dispensable for the function of U2 snRNP in splicing. *Genes Dev.*, **3**, 1887–1898.
Reddy,R. and Busch,H. (1988) Small nuclear RNAs: RNA sequences, structure, and modifications. In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp. 1–37.
Samarsky,D.A., Fournier,M.J., Singer,R.H. and Bertrand,E. (1998) The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J.*, **17**, 3747–3757.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Ségault,V., Will,C.L., Sproat,B.S. and Lührmann,R. (1995) *In vitro* reconstitution of mammalian U2 and U5 snRNPs active in splicing: Sm proteins are functionally interchangeable and are essential for the formation of functional U2 and U5 snRNPs. *EMBO J.*, **14**, 4010–4021.
Sleeman,J.E. and Lamond,A.I. (1999a) Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.*, **9**, 1065–1074.
Sleeman,J.E. and Lamond,A.I. (1999b) Nuclear organization of pre-mRNA splicing factors. *Curr. Opin. Cell Biol.*, **11**, 372–377.
Smith,K.P. and Lawrence,J.B. (2000) Interactions of U2 gene loci and their nuclear transcripts with Cajal (coiled) bodies: Evidence for preU2 within Cajal bodies. *Mol. Biol. Cell*, **11**, 2987–2998.
Smith,C.M. and Steitz,J.A. (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, **89**, 669–672.
Spector,D.L. (1993) Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.*, **9**, 265–315.
Tollervey,D. and Kiss,T. (1997) Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.*, **9**, 337–342.
Tollervey,D., Lehtonen,H., Jansen,R., Kern,H. and Hurt,E.C. (1993) Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell*, **72**, 443–457.
Tyc,K. and Steitz,J.A. (1989) U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J.*, **8**, 3113–3119.
Tycowski,K.T., You,Z.H., Graham,P.J. and Steitz,J.A. (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell*, **2**, 629–638.
Wang,H., Boisvert,D., Kim,K.K., Kim,R. and Kim,S.H. (2000) Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution. *EMBO J.*, **19**, 317–323.
Weinstein,L.B. and Steitz,J.A. (1999) Guided tours: from precursor snoRNA to functional snoRNP. *Curr. Opin. Cell Biol.*, **11**, 378–384.
Will,C.L. and Lührmann,R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.*, **13**, 290–301.
Yu,Y.T., Shu,M.D. and Steitz,J.A. (1998) Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.*, **17**, 5783–5795.
Yu,Y.T., Shu,M.D., Narayanan,A., Terns,R.M., Terns,M.P. and Steitz, J.A. (2001) Internal modification of U2 small nuclear (sn)RNA occurs in nucleoli of *Xenopus* oocytes. *J. Cell Biol.*, **152**, 1279–1288.
Zebarjadian,Y., King,T., Fournier,M.J., Clarke,L. and Carbon,J. (1999) Point mutations in yeast CBF5 can abolish *in vivo* pseudouridylation of rRNA. *Mol. Cell. Biol.*, **19**, 7461–7472.

Received February 25, 2002; revised March 18, 2002; accepted March 25, 2002

PUBLICATION II.

Kiss A.M., Jády B.E., Darzacq X., Verheggen C., Bertrand E. and Kiss T.

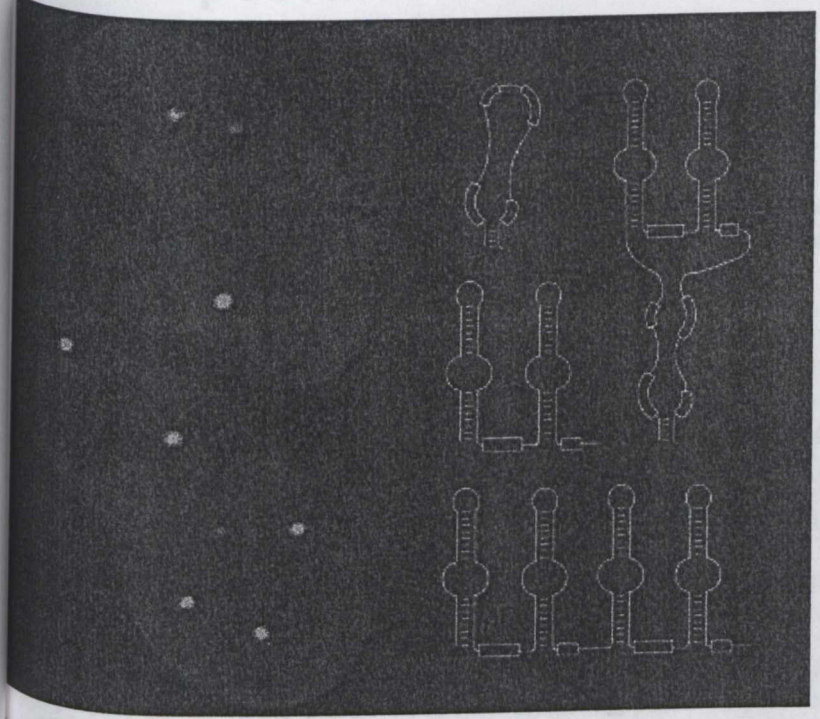
A CAJAL BODY-SPECIFIC PSEUDOURIDYLATION GUIDE RNA IS COMPOSED OF TWO BOX H/ACA SNORNA-LIKE DOMAINS.

Nucl Acid Res, **30**, (21) 4643-4649, 2002

INTRODUCTION

Several recently identified scaRNAs directing the 2'-*O*-ribose methylation and pseudouridylation of RNAPII-transcribed snRNAs were composed of a box C/D and box H/ACA snoRNA-like domains. During the sequence analysis of a cDNA library of human box H/ACA snRNAs, we have also identified a novel type of pseudouridylation guide RNA, the U93 RNA, that is composed of two tandemly arranged H/ACA snRNA domains and functions in pseudouridylation of the U2 spliceosomal snRNA. *In situ* hybridisation revealed a Cajal body-specific localization for U93, further supporting the idea that Cajal body provides the cellular locale for the post-transcriptional modification of spliceosomal snRNAs. By mutational analysis we demonstrated that all box elements of U*93* snRNA are necessary for stable accumulation of the full length RNA. However, both the 5'- and the 3'-terminal box H/ACA RNA domains of U*93* RNA could independently accumulate in the cell nucleus and posses functionally active localisation signals that can direct them into Cajal bodies.

About the Cover



over. Small Cajal body-specific RNAs (scaRNAs). Structural classification of a recently discovered class of 2′*O*-bose-methylation and pseudouridylation guide RNAs which reside in the nucleoplasmic Cajal (coiled) body and direct odification of the RNA polymerase II-transcribed U1, U2, U4 and U5 spliceosomal snRNAs. The U93 scaRNA co-calises with p80 coilin, a molecular marker of the Cajal body. For further details see the paper by Kiss *et al.* in this sue [*Nucleic Acids Res.* (2002) **30**, 4643--4649].

Table of Contents]

OME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH

Nucleic Acids Research, 2002, Vol. 30 No. 21 4643–4649

A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains

Arnold M. Kiss[1,2], Beáta E. Jády[3], Xavier Darzacq[1], Céline Verheggen[3], Edouard Bertrand[3] and Tamás Kiss[1,2,*]

[1]Laboratoire de Biologie Moléculaire Eucaryote du CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France, [2]Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary and [3]Institut de Génétique Moléculaire, Montpellier, France

Received July 17, 2002; Revised and Accepted September 9, 2002

DDBJ/EMBL/GenBank accession no. AF492209

ABSTRACT

Site-specific post-transcriptional conversion of uridines to pseudouridine in ribosomal RNAs and small nuclear RNAs (snRNAs) is directed by guide RNAs which possess the conserved box H and ACA sequence elements and fold into the consensus 'hairpin–hinge–hairpin–tail' secondary structure. Here, we describe an unusual mammalian pseudouridylation guide RNA, called U93, that is composed of two tandemly arranged box H/ACA RNA domains. The U93 RNA therefore carries two H and two ACA box motifs, all of which are essential for accumulation of the full-length RNA. The human U93 RNA accumulates in Cajal (coiled) bodies and it is predicted to function in pseudouridylation of the U2 spliceosomal snRNA. Our results lend further support to the notion that modification of the RNA polymerase II-transcribed spliceosomal snRNAs takes place in Cajal bodies.

INTRODUCTION

The nucleus contains a large number of small nuclear RNAs (snRNAs) which possess diverse cellular functions (1). The U1, U2, U4, U5 and U6 spliceosomal snRNAs play a pivotal role in the removal of intron regions from pre-mRNAs. The U7 snRNA directs 3′-end formation of histone mRNAs and the RNase P RNA functions in 5′-end processing of tRNAs. The U3, U8, U14, U22, snR30 and MRP small nucleolar RNAs (snoRNAs) are required for the production of mature ribosomal RNAs (rRNAs) (2). Besides RNA processing, snRNAs also function in the regulation of transcription elongation by RNA polymerase (pol) II (7SK RNA) (3,4) and in the synthesis of telomeric DNA repeats (telomerase RNA) (5). All snRNAs associate with specific proteins and form small nuclear or nucleolar ribonucleoprotein particles (snRNPs or snoRNPs) (1).

The nucleus also contains an enormous number of modification guide RNAs which direct the post-transcriptional synthesis of 2′-*O*-methylated nucleotides and pseudouridines in rRNAs, spliceosomal snRNAs and, most likely, other cellular RNAs (for recent reviews, see 6–11). The 2′-*O*-methylation guide RNAs possess the conserved box C (RUGAUGA) and D (CUGA) elements that are frequently tethered by a short 5′,3′-terminal stem (Fig. 1). The internal region of box C/D RNAs carries the C′ and D′ boxes which represent perfect or imperfect copies of the C and D boxes (12,13). The pseudouridylation guide RNAs are characterised by a consensus 'hairpin–hinge–hairpin–tail' structure and they carry the conserved H (ANANNA) and ACA box elements (14,15). The H box is located in the single-stranded hinge region after the 5′ hairpin. The ACA box is found in the 3′-terminal tail 3 nt before the end of the RNA. The box C/D 2′-*O*-methylation guide RNAs are associated with four RNP proteins, fibrillarin/Nop1p, Nop56p, Nop58p and 15.5kD/Snu13p, whereas the box H/ACA pseudouridylation guide RNAs specifically bind dyskerin/Cbf5p, Gar1p, Nhp2p and Nop10p (reviewed in 8–10). The guide RNAs provide scaffolding for the associated RNP proteins and select the target nucleotide by forming direct base-pairing interactions with the substrate RNA. In turn, the RNP proteins provide metabolic stability for the guide RNAs and catalyse the modification reactions. Most likely, the fibrillarin/Nop1p and dyskerin/Cbf5p RNP proteins perform the 2′-*O*-methyl transfer and the uridine to pseudouridine isomerisation reactions, respectively (16,17).

The modification guide RNAs accumulate either in the nucleolus or in the Cajal body and, accordingly, they are called snoRNAs or small Cajal body-specific RNAs (scaRNAs) (2,18). The two groups of modification guide RNAs possess different functions. The snoRNAs direct 2′-*O*-methylation and pseudouridylation of rRNAs (19–23) and the RNA pol III-transcribed U6 spliceosomal snRNA (24–26). The scaRNAs function in modification of the RNA pol II-specific U1, U2, U4 and U5 spliceosomal snRNAs (18,27).

*To whom correspondence should be addressed at Laboratoire de Biologie Moléculaire Eucaryote du CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France. Tel: +33 5 61 33 59 91; Fax: +33 5 61 33 58 86; Email: tamas@ibcg.biotoul.fr

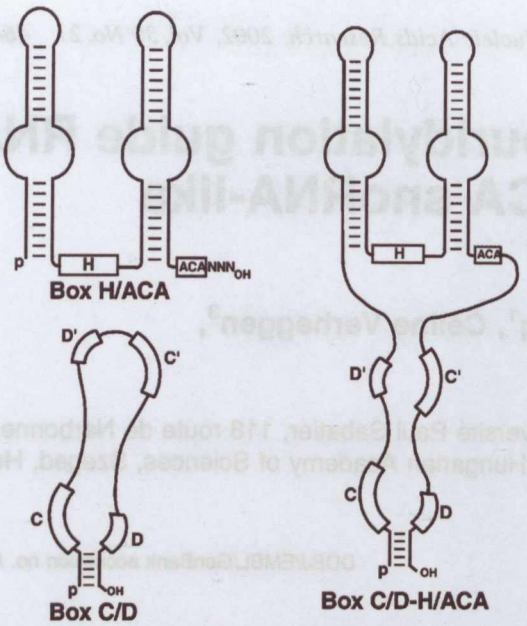


Figure 1. Classification of 2′-*O*-methylation and pseudouridylation guide RNAs. Schematic representation of the consensus two-dimensional structures of box C/D 2′-*O*-methylation, box H/ACA pseudouridylation and box C/D-H/ACA composite 2′-*O*-methylation and pseudouridylation guide RNAs. Positions of the conserved box elements are indicated.

The scaRNAs are frequently composed of a box C/D and a box H/ACA RNA-like domain (Fig. 1) and they can function both in 2′-*O*-methylation and pseudouridylation of snRNAs (18,27). Here, we describe a novel mammalian scaRNA, called the U93 RNA, that is predicted to function in pseudouridylation of the RNA pol II-specific U2 spliceosomal snRNA. Surprisingly, the human, mouse and cow U93 scaRNAs are composed of two tandemly repeated box H/ACA domains and therefore they contain two H and two ACA boxes which are all essential for accumulation of the full-length U93 RNA.

MATERIALS AND METHODS

General procedures

Unless stated otherwise, all techniques for manipulating DNA, RNA and oligonucleotides have been done as described by Sambrook *et al.* (28). The following oligonucleotides were used in this study: 1, ATGCCTCAGCTTCCTCT; 2, CAGACTTGCAGAAAAAGCA; 3, CAGTACTTAGTGTTCAACAGA; 4, ATAATCGATAATCTGTAGTCTTGGAGCCGC; 5, ATAGTCGACACTTGTGGCAGTACTTAGTG; 6, CAGCTTCCTCTGGAGTAGGGGGGTTCTGCAAGTCTGGTGT; 7, CAACAGTGACCAGAAACGGGCAGAGGAAAATTGCACA; 8, TCTATCCGCCAACAGTACCCCCCGCTTTGCAGTCGAGATA; 9, TATAGTCGACACTGGGGGCAGTACTTAGTGTTC; 10, GTGATAATGACTGGGCTATGTC; 11, AATACGACTCACTATAGGGGGCAGTACTTAGTGTTCAACAG; 12, AT*CTCGAGGGTACCACCAAT*CCTGTGCGGCTCCAT*A (amino-allyl-modified T residues are marked by asterisks).

Characterisation of human, mouse and cow U93 RNAs

A recombinant plasmid carrying a full-length cDNA of the human U93 scaRNA was identified during characterisation of a cDNA library of human HeLa snRNAs. Synthesis and cloning of cDNAs has been described (19). The 5′ terminus of U93 was confirmed by primer extension analysis using the terminally labelled oligonucleotide 1 as a primer. Partial cDNAs of the mouse and cow U93 scaRNAs were obtained by RT–PCR using oligonucleotides 2 and 3 as primers. The PCR-amplified fragments were cloned into the *Sma*I site of pBluescript (Stratagene) and subjected to sequence analyses. The 3′-terminal sequences of the mouse and cow U93 RNAs were determined by the oligoribonucleotide ligation–PCR amplification procedure (29), except that oligonucleotide 3 was used as a U93-specific upstream primer. The 5′-terminal sequences of mouse and cow U93 RNAs were determined by the RNA 5′-RACE procedure using oligonucleotide 1 as a primer.

Expression constructs

The coding region of the human U93 RNA gene was PCR amplified by using oligonucleotides 4 and 5 as primers and human genomic DNA as a template. The PCR product was digested with *Cla*I and *Sal*I and inserted into the *Cla*I and *Xho*I sites of the pCMV-globin expression vector (18), resulting in pCMV-globin-U93. Construction of derivatives of pCMV-globin-U93 carrying U93 genes with altered H1, H2 and ACA1 boxes was performed by the megaprimer amplification approach using pCMV-globin-H93 as a template (30). To obtain pCMV-globin-U93*H1* and pCMV-globin-U93*ACA1* the 5′ half of the U93 gene was amplified with oligonucleotides 6 and 7 as mutagenic 3′ primers, respectively, and oligonucleotide 4 as a 5′-end-specific primer. The PCR products were used as megaprimers in the second amplification step in combination with a common 3′-end-specific primer (oligonucleotide 5). The amplified products were digested with *Cla*I and *Mlu*I and cloned into the *Cla*I and *Mlu*I sites of pCMV-globin. A similar strategy was used to generate pCMV-globin-U93*H2*, except that a megaprimer was generated with oligonucleotides 5 and 8 as 3′- and 5′-end-specific primers, respectively. In the second step of amplification oligonucleotide 4 was used as a 5′-end-specific primer. Finally, amplification of the human U93 RNA gene with oligonucleotides 4 and 9 resulted in U93*ACA2*, which was inserted into the *Cla*I and *Xho*I sites of pCMV-globin, yielding pCMV-globin-U93*ACA2*. The identity of each expression construct was verified by sequence analyses. Construction of the coilin-Dsred2 expression vector has been described (31). Transfection of simian COS-7, human HeLa and mouse L929 cells has been described (18).

Immunoprecipitation, cell fractionation and RNA analysis

Immunoprecipitation of human snRNPs with anti-fibrillarin and anti-hGAR1 antibodies (kindly provided by Drs J. A. Steitz and W. Filipowicz, respectively) was performed as described (18). Fractionation of HeLa cells into nuclear, nucleoplasmic, nucleolar and cytoplasmic fractions was performed according to Tyc and Steitz (32). From COS-7 and HeLa cells and the nuclear, nucleolar, nucleoplasmic and

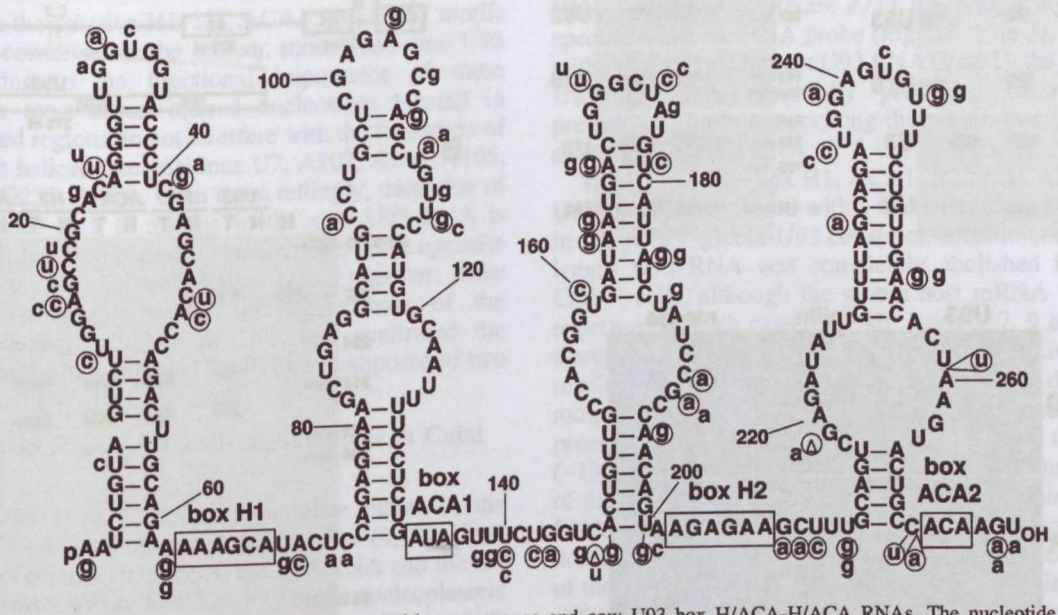


Figure 2. Primary and predicted two-dimensional structures of human, mouse and cow U93 box H/ACA-H/ACA RNAs. The nucleotide sequence of the human U93 RNA is in capitals. Lower case and circled lower case letters indicate changes in the mouse (*Mus musculus*) and cow (*Bos taurus*) U93 RNAs, respectively. Nucleotide deletions (Δ) and insertions (arrowheads) are also indicated. The putative box H (H1 and H2) and ACA (ACA1 and ACA2) sequences are boxed. Please note that nucleotides which might be involved in formation of putative pseudouridylation pockets are unfolded. The sequence of the human U93 RNAs is available in GenBank under the accession no. AF492209.

cytoplasmic fractions of HeLa cells, RNA was isolated by the guanidine thiocyanate/phenol–chloroform extraction procedure (33). RNase A/T1 protection was performed as described by Goodall *et al.* (33). For mapping of U93, U93*H1*, U93*ACA1*, U93*H2* and U93*ACA2* RNAs, sequence-specific RNA probes were synthesised by SP6 RNA pol using *Nde*I-digested pCMV-globin-U93, pCMV-globin-U93*H1*, pCMV-globin-U93*ACA1*, pCMV-globin-U93*H2* and pCMV-globin-U93*ACA2* plasmids as templates. Probes for U3, U4 and U19 snRNAs have been described (18). After synthesis, all probes were purified on a 6% sequencing gel.

In situ hybridisation

Fluorescent *in situ* hybridisation, image acquisition and processing has been described (18). To generate antisense RNA probe for the human U93 RNA, a fragment of the U93 gene from position 159 to position 269 was PCR-amplified with oligonucleotides 10 and 11 as primers. Utilisation of oligonucleotide 11 as a 3′-end-specific primer resulted in inclusion of the T7 RNA polymerase promoter into the amplified fragment. The resulting PCR product was used as a template for *in vitro* transcription by T7 RNA polymerase in the presence of 5-(3-aminoallyl) uridine 5′-triphosphate. Detection of the 5′-terminal box H/ACA domain of the U93 RNA was performed by using an oligonucleotide probe complementary to the human U93 RNA from position 13 to 44 (oligonucleotide 12). The modified RNA and oligonucleotide probes were labelled with FluoroLink™ Cy5-monofunctional dye (Amersham) according to the protocol of the laboratory of Dr Singer (http://singerlab.aecom.yu.edu). Human p80 coilin has been detected by polyclonal rabbit anti-coilin antibody (kindly provided by Dr A. Lamond) as it has been described (18).

RESULTS

Identification of a novel Gar1p-associated human box H/ACA RNA

During sequence analysis of a cDNA library of human HeLa snRNAs, we have identified a 275 nt long RNA that showed no significant sequence similarity to any known human RNA (Fig. 2). Northern analysis and RNase A/T1 protection experiments confirmed that the new RNA, called hereafter as U93, efficiently accumulates in HeLa cells (Figs 3 and 4; data not shown). The presence of an ACA triplet 3 nt before the 3′ terminus of the U93 RNA indicated that it might belong to the family of box H/ACA RNAs. Indeed, a monoclonal antibody directed against the human Gar1 protein, a component of box H/ACA RNPs, specifically precipitated the U93 RNA as well as the U19 box H/ACA snoRNA from a human HeLa cell extract (Fig. 3A). In contrast, an anti-fibrillarin antibody failed to precipitate both U93 and U19, but recognised the fibrillarin-associated U3 box C/D snoRNA. As expected, neither the anti-fibrillarin nor the anti-GAR1 antibody reacted with the U4 spliceosomal snRNP, demonstrating that the human U93 RNA specifically associates with the Gar1 snoRNP protein and it belongs to the family of box H/ACA RNAs.

Mammalian U93 snRNA is composed of two box H/ACA RNA-like domains

Computer modelling was used to generate a secondary structure for the human U93 RNA (34). In contrast to the consensus 'hairpin–hinge–hairpin–tail' structure of box H/ACA snoRNAs (Fig. 1), the computer-predicted structure of U93 is composed of four hairpins connected by three single-stranded hinge regions and terminated by a short tail (Fig. 2).

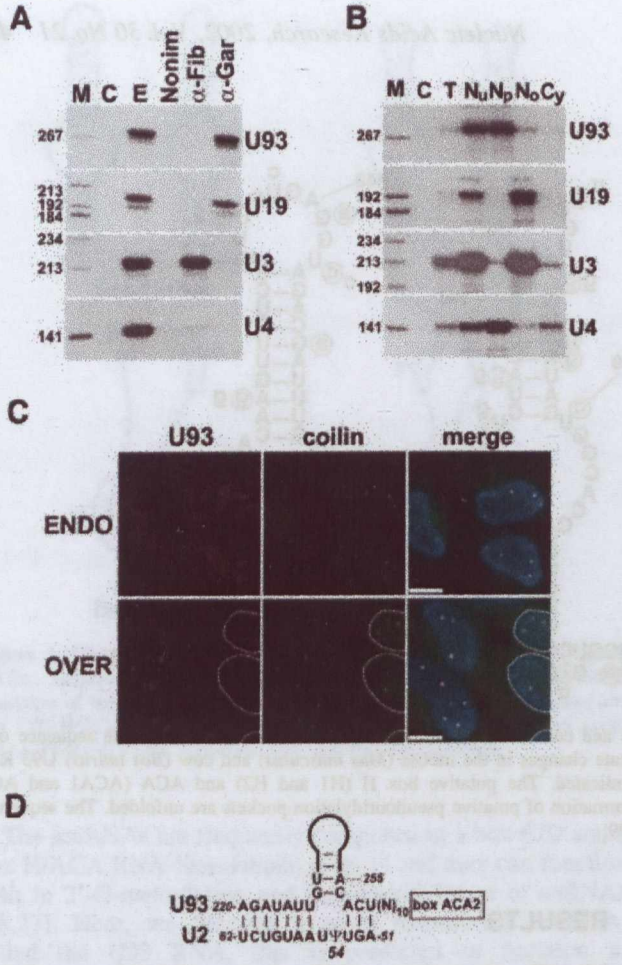


Figure 3. Characterisation of human U93 RNA. (**A**) Immunological characterisation of U93 RNP. RNAs recovered from the pellet of immunoprecipitation reactions performed with anti-hGAR1 (α-Gar), anti-fibrillarin (α-Fib) or a control non-immune serum (Nonim) from a human HeLa cell extract were mapped with sequence-specific antisense RNA probes indicated on the right. Lanes C and E show control mappings with RNAs obtained from *Escherichia coli* and HeLa cell extracts, respectively. Lane M, size marker (a mixture of *Hea*III- and *Taq*I-digested pBR322). (**B**) Cell fractionation. RNA isolated from HeLa cells (T), from the nuclear (Nu), nucleoplasmic (Np), nucleolar (No) and cytoplasmic (Cy) fractions of HeLa cells were mapped by sequence-specific antisense RNA probes as indicated on the right. (**C**) *In situ* localisation of U93 RNA. Human HeLa cells either transfected (bottom) or non-transfected (top) with the pCMV-globin-U93 expression construct (see Fig. 4A) were probed with a fluorescent RNA probe complementary to the human U93 RNA. Cajal bodies were detected by an antibody directed against human p80 coilin. Merged images show that the U93 RNA co-localises with p80 coilin both in transfected and non-transfected cells. The nuclei of non-transfected cells are highlighted by dotted lines. Under the exposure conditions shown, the endogenous U93 RNA remains invisible in non-transfected cells. (**D**) Potential base-pairing interaction of the human U93 and U2 RNAs. The 3′-terminal hairpin of U93 is schematically indicated. The U54 residue known to be pseudouridylated in vertebrate U2 snRNAs is indicated (Ψ).

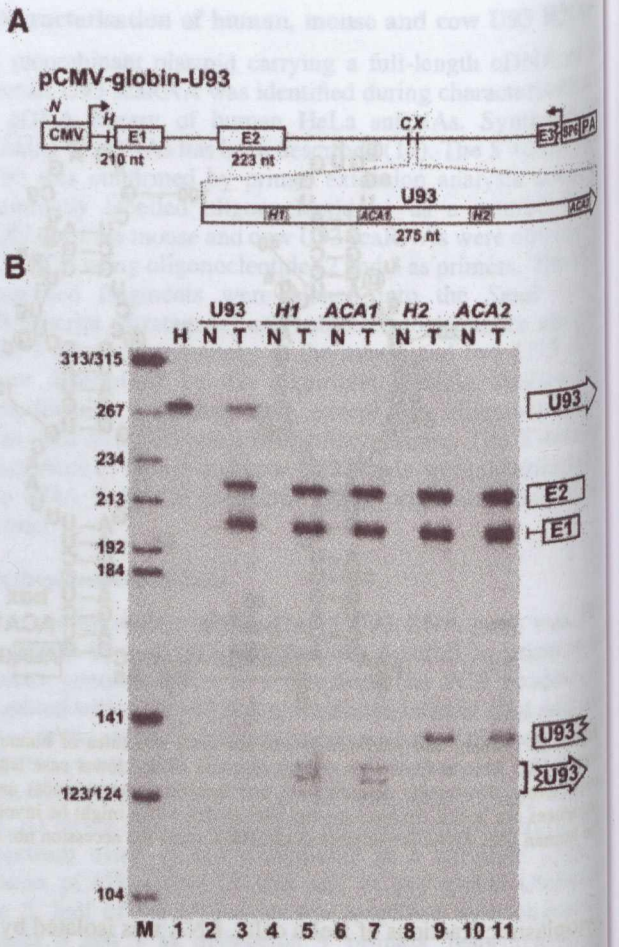


Figure 4. Transient expression of human U93 RNA in COS-7 cells. (**A**) Schematic representation of the pCMV-globin expression construct. The CMV promoter, full-length (E1 and E2) and partial (E3) exons of the human β-globin gene, the SP6 RNA polymerase promoter (SP6) and the polyadenylation region of the bovine growth hormone gene (PA) are indicated. Relevant restriction sites are shown (*N*, *Nde*I; *H*, *Hin*dIII; *C*, *Cla*I; and *X*, *Xho*I). Open arrow indicates the coding region of human U93 RNA. In the U93-*H1*, U93-*ACA1*, U93-*H2* and U93-*ACA2* constructs, the indicated box element was replaced for a stretch of C residues. (**B**) RNase A/T1 protection analysis. Plasmids used for transfection of COS-7 cells and specific probes utilised for mappings are indicated at the top. Lanes N and T represent mapping reactions with RNAs obtained from non-transfected and transfected cells, respectively. Lane H, mapping of HeLa cellular RNAs. RNA fragments protected by the first and second exons of the spliced β-globin mRNA, the U93 RNA and the 5′- and 3′-terminal portions of the U93 RNA are indicated on the right. Lane M, size marker in nucleotides.

We noticed that the first and third hairpins of U93 are followed by potential box H sequences (63-AAAGCA-68 and 202-AGAGAA-207) and that the second and the last hairpins are followed by potential box ACA motifs (134-AUA-136 and 270-ACA-272). This strongly suggested that the human U93 RNA is composed of two tandemly arranged box H/ACA RNA domains.

Northern analysis of mouse, rat, cow and carp nuclear RNAs performed with a human U93-specific antisense RNA probe revealed hybridising RNAs of the size of human U93 (data not shown). The full-length sequences of the mouse and cow U93 RNAs were obtained by RT–PCR followed by 5′- and 3′-end RACE experiments (see Materials and Methods). A comparison of the sequences and two-dimensional structures of the human, mouse and cow U93 RNAs is shown in Figure 5. Strikingly, most of the evolutionarily variable nucleotides are located in predicted single-stranded regions, namely in the three hinge and the 3′-terminal tail regions as well as in internal and terminal loop structures. Contrary to the fact that the hinge regions show the highest sequence variability

nucleotides in the putative H1, H2, ACA1 and ACA2 motifs are perfectly conserved in the human, mouse and cow U93 RNAs, pointing to the functional importance of these sequences. In most cases, altered nucleotides located in double-stranded regions do not interfere with the formation of perfect double helices (see positions U7, A107, A164, A165, A196, G184, G232, A254). Even more tellingly, transition of the A168 residue to a G residue in the cow U93 RNA is restored by a U to C compensatory base change in the opposite strand of the proposed double helix. Taken together, these observations strongly supported the correctness of the proposed secondary structure of U93 and confirmed the conclusion that the mammalian U93 RNA is composed of two box H/ACA snoRNA-like domains.

The human U93 RNA specifically accumulates in Cajal bodies

Mammalian box H/ACA RNAs accumulate either in the nucleolus or in the nucleoplasmic Cajal bodies (8,18). Upon fractionation of human HeLa cells, the U93 RNA and the U4 spliceosomal snRNA were found mainly in the nucleoplasmic fraction, while the U19 box H/ACA and the U3 box C/D authentic snoRNAs co-purified with the nucleolar fraction (Fig. 3B). To determine the precise nucleoplasmic localisation of U93, we used *in situ* fluorescent microscopy (Fig. 3C). Probing of HeLa cells with a U93-specific fluorescent antisense RNA probe revealed that the U93 RNA localises to sharp dot-like structures in the nucleoplasm. Staining the same cells with an antibody directed against p80 coilin, a molecular marker of Cajal bodies (35), demonstrated that the U93 RNA perfectly co-localises with p80 coilin. Likewise, upon transient overexpression of the human U93 RNA in HeLa cells (for details, see below), the overproduced U93 RNA showed a specific co-localisation with p80 coilin. Thus far, all box C/D and box H/ACA RNAs residing in Cajal bodies have been implicated in post-transcriptional modification of pol II-specific spliceosomal snRNAs (8,11,18). Indeed, we noticed that sequences in the internal loop of the 3′-terminal hairpin of U93 are capable of positioning the U54 residue in the U2 snRNA for pseudouridylation (Fig. 3D). We concluded that (i) the U93 box H/ACA-H/ACA composite RNA represents a new member of scaRNAs and (ii) the U93 scaRNA likely functions in pseudouridylation of the RNA pol II-specific U2 spliceosomal snRNA.

Elements essential for accumulation of the U93 box H/ACA-H/ACA scaRNA

The box H and ACA elements play a crucial role in accumulation of yeast and mammalian box H/ACA snoRNAs (14,15). Since the U93 RNA possesses two putative box H (H1 and H2) and ACA (ACA1 and ACA2) elements, we have tested the functional importance of each of these elements in RNA accumulation. The genomic copy of the human U93 RNA is located in an intron of a series of reported spliced expressed sequence tags, indicating that the U93 RNA is generated by intron processing (36,37). Therefore, to express the human U93 RNA, its coding region was inserted into the second intron of the human β-globin gene controlled by the cytomegalovirus (CMV) promoter (Fig. 4A). The resulting pCMV-globin-U93 expression construct was transfected into simian COS-7 cells and expression of the U93 RNA was tested by RNase A/T1 mapping using a sequence-specific antisense RNA probe (Fig. 4B, lane 3). As compared with the authentic human U93 RNA (lane 1), the plasmid-born U93 RNA was faithfully processed from the globin pre-mRNA, further supporting the notion that U93 is indeed an intron-encoded RNA.

When either the box H1, ACA1, H2 or ACA2 motif of the U93 gene was replaced with a C stretch of appropriate length in the pCMV-globin-U93 construct, accumulation of the full-length U93 RNA was completely abolished in transfected COS-7 cells, although the globin host mRNA was correctly expressed in each case (Fig. 4B, lanes 5, 7, 9 and 11). Upon disruption of the box H1 or ACA1 motif, RNase mapping resulted in accumulation of a protected RNA doublet with a molecular size that corresponds to the predicted length of the processed 3′-terminal box H/ACA domain of the U93 RNA (~130 nt) (Fig. 4B, lanes 5 and 7). On the other hand, alteration of the H2 or ACA2 motif resulted in the accumulation of an RNA of the expected size (~140 nt) of the 5′-terminal H/ACA domain of U93 (Fig. 4B, lanes 9 and 11). Indeed, the identity of these RNAs was confirmed by RNase mappings performed with probes specific for either the 5′- or the 3′-terminal half of the U93 RNA (data not shown). These results demonstrated that both the 5′- and the 3′-terminal box H/ACA RNA domains of the human U93 RNA can independently accumulate. However, mapping of HeLa cellular RNA failed to detect any accumulating fragment of the U93 RNA (Fig. 4B, lane 1, and data not shown), indicating that normally neither the 5′- nor the 3′-terminal box H/ACA domain of U93 is expressed separately.

We next investigated whether the 5′- and 3′-terminal domains of the U93 RNA, when they are expressed independently, still accumulate in Cajal bodies. Mouse cells transfected with the pCMV-globin expression construct carrying the U93, U93-*H1* or U93-*H2* genes were probed with fluorescent antisense probes specific for the 5′- and 3′-terminal domains of the U93 RNA (Fig. 5). The Cajal bodies of the transfected mouse cells were visualised by co-expression of the coilin-Dsred2 fluorescent recombinant protein. Similar to the full-length human U93 RNA, both the 5′- and 3′-terminal box H/ACA domains of the RNA accumulated in Cajal bodies. This demonstrates that both box H/ACA domains of U93 possess functionally active localisation signals which can direct the Cajal body-specific accumulation of the two domains independently of one another.

DISCUSSION

Site-specific synthesis of 2′-*O*-methylated nucleotides and pseudouridines in rRNAs and spliceosomal snRNAs is directed by guide RNAs which can be classified into three structurally and functionally well defined families (Fig. 1). The box C/D RNAs direct 2′-*O*-methylation and the box H/ACA RNAs guide pseudouridylation of rRNAs and snRNAs. Another minor group of composite box C/D-H/ACA RNAs function both in 2′-*O*-methylation and pseudouridylation of snRNAs (18,27). In this study, we have identified and characterised a novel type of pseudouridylation guide RNA. The U93 RNA is composed of two tandemly arranged box H/ACA RNA domains and, therefore, it carries two box H and ACA motifs. This unique architecture of the

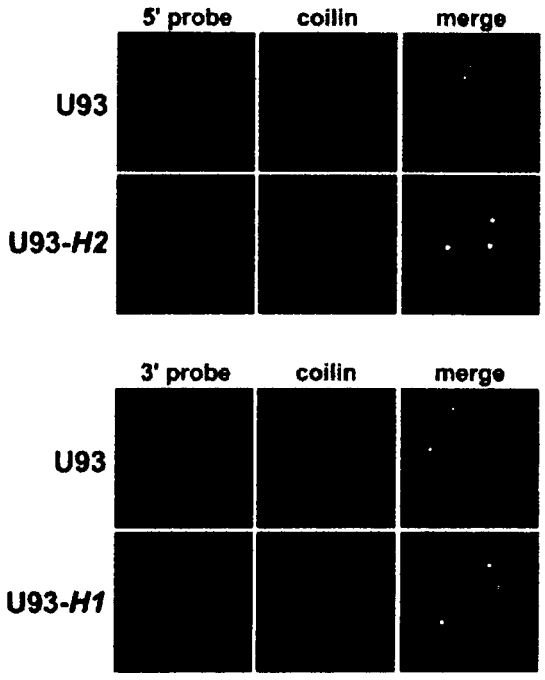


Figure 5. *In situ* localisation of the 5′- and 3′-terminal box H/ACA domains of human U93 expressed in mouse cells. Mouse L929 cells were transfected with the pCMV-globin-U93, pCMV-globin-U93*H1* and pCMV-globin-U93*H2* expression constructs (see Fig. 4A). The transiently expressed U93 RNA was visualised by a fluorescent oligonucleotide or an antisense RNA probe specific for the 5′- and 3′-terminal domain of the human U93 RNA, respectively. Cajal bodies of mouse cells were detected by co-expression of the coilin-Dsred2 fluorescent protein. For other details, see the legend to Figure 3C.

U93 RNA is conserved at least in human, mouse and cow (Fig. 2).

Previous characterisation of the human U85 box C/D-H/ACA composite RNA showed that the unusual structural organisation of this class of modification guide RNAs could be explained by several facts. Alteration of the 5′-terminal C or the 3′-terminal D box of the U85 RNA completely abolished the expression of both the full-length U85 RNA and box H/ACA domain of the RNA (27). This indicated that accumulation of the human U85 box C/D-H/ACA RNA is supported exclusively by its box C/D domain and that the box H/ACA domain of the RNA lacks metabolic stability. In the case of the newly discovered U93 box H/ACA-H/ACA composite RNA, all conserved box elements (H1, ACA1, H2 and ACA2) are required for accumulation of the full-length RNA. However, accumulation of the 5′- and 3′-terminal box H/ACA domains of U93 is not interdependent. In other words, both box H/ACA domains of the U93 RNA possess all those elements which are essential for accumulation (Fig. 4).

Thus far, all 2′-*O*-methylation and pseudouridylation guide RNAs implicated in modification of the RNA pol II-specific U1, U2, U4 and U5 spliceosomal snRNAs have been found to reside in the nucleoplasmic Cajal bodies (18). It seems that these guide RNAs possess specific *cis*-acting targeting elements which are responsible for the their Cajal body specific localisation. Currently, we have found that the Cajal body-specific localisation signals of the human U85 C/D-H/ACA scaRNA are located in the box H/ACA domain of the RNA and that the box C/D domain of U85 alone cannot localise to Cajal bodies (P.Richard, X.Darzacq, C.Verheggen, E.Bertrand and T.Kiss, manuscript in preparation). The human U93 box H/ACA-H/ACA RNA also accumulates in Cajal bodies and, unexpectedly, both box H/ACA domains of the U93 RNA possess functionally active Cajal body-specific targeting elements (Fig. 5). So, in contrast to the U85 box C/D-H/ACA RNA, the unusual structural organisation of the U93 box H/ACA-H/ACA RNA cannot be explained by the metabolic instability or incorrect intracellular trafficking of the two domains of the RNA.

The box C/D and H/ACA domains of the U85 RNA function in 2′-*O*-methylation and pseudouridylation of two neighbouring nucleotides in the U5 spliceosomal snRNA. This suggests that co-expression of guide RNAs directing modification of the same substrate RNA is advantageous for the cell (27). The putative pseudouridylation pocket in the 3′-terminal hairpin of the U93 RNA is predicted to direct pseudouridylation of the U2 spliceosomal snRNA at position U54 (Fig. 3D). Since both the 5′- and 3′-terminal hairpins of box H/ACA RNAs can carry functional pseudouridylation pockets (23,38), the four hairpins of the U93 box H/ACA-H/ACA RNA, in principle, could direct pseudouridylation of four different substrate uridines. It seems, however, unlikely that the first and third hairpins of U93 contain functional pseudouridylation pockets, since several evolutionarily altered nucleotides are present in these regions (Fig. 2). On the contrary, the second hairpin carries a conserved internal loop that possesses the characteristic topology of a functional pseudouridylation pocket. This putative pseudouridylation pocket of U93 lacks complementarity to rRNAs, spliceosomal snRNAs or any known stable RNAs, including tRNAs. Therefore, it seems likely that the second hairpin of the U93 RNA directs pseudouridylation of a not yet identified RNA. Whether this putative target RNA of U83, like all substrate RNAs of Cajal body-specific guide RNAs, has a cellular function related to pre-mRNA splicing remains an intriguing speculation.

In summary, identification of the U93 RNA that is composed of two box H/ACA RNA domains further supports the idea that snRNA structural domains are frequently used to build novel composite RNAs. In addition to the recently characterised box C/D-H/ACA scaRNAs (18,27), the vertebrate and yeast telomerase RNAs carry a box H/ACA and Sm domain, respectively (39,40). In principle, acquisition of structurally and functionally well defined snRNA domain may contribute to metabolic stability, may direct intracellular trafficking or may provide a new cellular function for the newly made composite RNA.

ACKNOWLEDGEMENTS

We are indebted to Y. de Preval for synthesis of oligonucleotides. A.M.K. was funded by the French Government and the Hungarian Academy of Sciences. B.E.J. is a recipient of a long term EMBO fellowship. X.D. and C.V. were funded by la Fondation pour la Recherche Médicale and Association pour la Recherche contre le Cancer, respectively. Our work was supported by grants from Association pour la Recherch

contre le Cancer, la Ligue Nationale contre le Cancer, the French MNRT (ACI) and the Hungarian Research Foundation (OTKA, T29042 and T31738).

REFERENCES

1. Yu,Y.T., Scharl,E.C., Smith,C.M. and Steitz,J.A. (1999) The growing world of small nuclear ribonucleoproteins. In Gesteland,R.F. Cech,T.R. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 487–524.
2. Maxwell,E.S. and Fournier,M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **64**, 897–934.
3. Nguyen,V.T., Kiss,T., Michels,A.A. and Bensaude,O. (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature*, **414**, 322–325.
4. Yang,Z., Zhu,Q., Luo,K. and Zhou,Q. (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature*, **414**, 317–322.
5. Blackburn,E. (1999) Telomerase. In Gesteland,R.F, Cech,T.R. and Atkins,J.F. (eds.), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 609–635.
6. Smith,C.M. and Steitz,J.A. (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, **89**, 669–672.
7. Bachellerie,J.P., Cavaillé,J. and Qu,L.H. (2000) Nucleotide modifications of eukaryotic rRNAs: the world of small nucleolar RNAs revisited. In Garrett,R.A., Douthwaite,S.R., Matheson,A.T., Moore,P.B. and Noller,H.F. (eds), *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions*. ASM Press, Washington, pp. 191–203.
8. Kiss,T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**, 3617–3622.
9. Filipowicz,W. and Pogacic,V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.*, **14**, 319–327.
10. Terns,M.P. and Terns,R.M. (2002) Small nucleolar RNAs: versatile *trans*-acting molecules of ancient evolutionary origin. *Gene Expr.*, **10**, 17–39.
11. Kiss,T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
12. Tycowski,K.T., Shu,M.D. and Steitz,J.A. (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, **379**, 464–466.
13. Kiss-László,Z., Henry,Y. and Kiss,T. (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.*, **17**, 797–807.
14. Balakin,A.G., Smith,L. and Fournier,M.J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, **86**, 823–834.
15. Ganot,P., Caizergues-Ferrer,M. and Kiss,T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**, 941–956.
16. Omer,A.D., Ziesche,S., Ebhardt,H. and Dennis,P.P. (2002) In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Natl Acad. Sci. USA*, **99**, 5289–5294.
17. Hoang,C. and Ferre-D'Amare,A.R. (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929–939.
18. Darzacq,X., Jády,B.E., Verheggen,C., Kiss,A.M., Bertrand,E. and Kiss,T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2′-*O*-methylation and pseudouridylation guide RNAs. *EMBO J.*, **21**, 2746–2756.
19. Kiss-László,Z., Henry,Y., Bachellerie,J.P., Caizergues-Ferrer,M. and Kiss,T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*, **85**, 1077–1088.
20. Cavaillé,J., Nicoloso,M. and Bachellerie,J.P. (1996) Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature*, **383**, 732–735.
21. Tycowski,K.T., Smith,C.M., Shu,M.D. and Steitz,J.A. (1996) A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in *Xenopus*. *Proc. Natl Acad. Sci. USA*, **93**, 14480–14485.
22. Ni,J., Tien,A.L. and Fournier,M.J. (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*, **89**, 565–573.
23. Ganot,P., Bortolin,M.L. and Kiss,T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.
24. Tycowski,K.T., You,Z.H., Graham,P.J. and Steitz,J.A. (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell*, **2**, 629–638.
25. Ganot,P., Jády,B.E., Bortolin,M.L., Darzacq,X. and Kiss,T. (1999) Nucleolar factors direct the 2′-*O*-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol. Cell. Biol.*, **19**, 6906–6917.
26. Lange,T.S. and Gerbi,S.A. (2000) Transient nucleolar localization Of U6 small nuclear RNA in *Xenopus laevis* oocytes. *Mol. Biol. Cell*, **11**, 2419–2428.
27. Jády,B.E. and Kiss,T. (2001) A small nucleolar guide RNA functions both in 2′-*O*-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J.*, **20**, 541–551.
28. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
29. Kiss,T. and Filipowicz,W. (1993) Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. *EMBO J.*, **12**, 2913–2920.
30. Datta,A.K. (1995) Efficient amplification using 'megaprimer' by asymmetric polymerase chain reaction. *Nucleic Acids Res.*, **23**, 4530–4531.
31. Boulon,S., Basyuk,E., Blanchard,J.M., Bertrand,E. and Verheggen,C. (2002) Intra-nuclear trafficking: insight from live cell imaging. *Biochimie*, in press.
32. Tyc,K. and Steitz,J.A. (1989) U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J.*, **8**, 3113–3119.
33. Goodall,G.J., Wiebauer,K. and Filipowicz,W. (1990) Analysis of pre-mRNA processing in transfected plant protoplasts. In Dahlberg,J.E. and Abelson,J.N. (eds), *Methods in Enzymology*, Academic Press, New York, Vol. 181, pp. 148–161.
34. Zuker,M., Mathews,D.H. and Turner,D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In Barciszewski,J. and Clark,B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 11–43.
35. Andrade,L.E., Chan,E.K., Raska,I., Peebles,C.L., Roos,G. and Tan,E.M. (1991) Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. *J. Exp. Med.*, **173**, 1407–1419.
36. Tollervey,D. and Kiss,T. (1997) Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.*, **9**, 337–342.
37. Weinstein,L.B. and Steitz,J.A. (1999) Guided tours: from precursor snoRNA to functional snoRNP. *Curr. Opin. Cell Biol.*, **11**, 378–384.
38. Bortolin,M.L., Ganot,P. and Kiss,T. (1999) Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J.*, **18**, 457–469.
39. Seto,A.G., Zaug,A.J., Sobel,S.G., Wolin,S.L. and Cech,T.R. (1999) *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. *Nature*, **401**, 177–180.
40. Mitchell,J.R., Cheng,J. and Collins,K. (1999) A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3′ end. *Mol. Cell. Biol.*, **19**, 567–576.

PUBLICATION III.

Kiss A.M., Jády B.E., Bertrand E. and Kiss T.

THE HUMAN BOX H/ACA PSEUDOURIDYLATION GUIDE RNA MACHINERY.

Mol Cell Biol., **24**, 5797-5807, 2004

INTRODUCTION

Pseudouridine, the most abundant modified nucleoside in stable cellular RNAs is synthesised by posttranscriptional isomerisation of uridines. In eukaryotic RNAs, site-specific synthesis of pseudouridines is directed primarily by box H/ACA guide RNAs. Moreover, box H/ACA RNA domains are present in several important cellular RNAs that apparently do not function in pseudouridylation. Since only a limited number of box H/ACA RNAs were known in humans, it remained unclear to what extent box H/ACA domain containing RNAs participate in different cellular processes. By systematic search for novel box H/ACA snRNAs we identified 61 previously unknown RNAs. As expected most of the newly identified box H/ACAs are predicted to direct pseudouridylation of ribosomal RNAs and spliceosomal snRNAs. The human 18S, 5.8S and 28S rRNA contain almost 100 pseudouridines. Currently, we can link guide RNAs to more then 80% of the known ribosomal and spliceosomal pseudouridylation sites. We also identified the first human guide RNA predicted to guide pseudouridylation of the RNAPIII-specific U6 spliceosomal RNA and several scaRNAs, involved in pseudouridylation of RNAPII-specific snRNAs. More importantly, we have identified 12 novel box H/ACA snRNAs of unknown function. Also the genomic organization of box H/ACA snRNAs has been analysed. I also identified two novel 28S rRNA methylation guide box C/D snoRNAs, which were genetically linked to box H/ACA snoRNAs.

MOLECULAR AND CELLULAR BIOLOGY, July 2004, p. 5797–5807 Vol. 24, No. 13
0270-7306/04/$08.00+0 DOI: 10.1128/MCB.24.13.5797–5807.2004

Human Box H/ACA Pseudouridylation Guide RNA Machinery†

Arnold M. Kiss,[1,2] Beáta E. Jády,[2,3] Edouard Bertrand,[3] and Tamás Kiss[1,2]*

Laboratoire de Biologie Moléculaire Eucaryote du CNRS, UMR5099, IFR109 CNRS, 31062 Toulouse Cedex 4,[1] and Institut de Génétique Moléculaire, 34000 Montpellier,[3] France, and Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary[2]

Received 26 January 2004/Returned for modification 23 March 2004/Accepted 1 April 2004

Pseudouridine, the most abundant modified nucleoside in RNA, is synthesized by posttranscriptional isomerization of uridines. In eukaryotic RNAs, site-specific synthesis of pseudouridines is directed primarily by box H/ACA guide RNAs. In this study, we have identified 61 novel putative pseudouridylation guide RNAs by construction and characterization of a cDNA library of human box H/ACA RNAs. The majority of the new box H/ACA RNAs are predicted to direct pseudouridine synthesis in rRNAs and spliceosomal small nuclear RNAs. We can attribute RNA-directed modification to 79 of the 97 pseudouridylation sites present in the human 18S, 5.8S, and 28S rRNAs and to 11 of the 21 pseudouridines reported for the U1, U2, U4, U5, and U6 spliceosomal RNAs. We have also identified 12 novel box H/ACA RNAs which lack apparent target pseudouridines in rRNAs and small nuclear RNAs. These putative guide RNAs likely function in the pseudouridylation of some other types of cellular RNAs, suggesting that RNA-guided pseudouridylation is more general than assumed before. The genomic organization of the new box H/ACA RNA genes indicates that in human cells, all box H/ACA pseudouridylation guide RNAs are processed from introns of pre-mRNA transcripts which either encode a protein product or lack protein-coding capacity.

Posttranscriptional covalent modification of ribonucleotides is an important step in the biosynthesis of stable cellular RNAs, including tRNAs, rRNAs, small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs) (40). Biochemical, biophysical, and genetic studies have shown that modified nucleotides are important for the appropriate function of mature RNAs; they facilitate correct RNA folding and contribute to the formation of appropriate RNA-RNA and RNA-protein interactions (reviewed in references 1, 7, 12, 15, and 44).

While in tRNAs, most modified nucleotides are synthesized by protein enzymes, in eukaryotic rRNAs and snRNAs, site-specific synthesis of the most prevalent modified ribonucleotides, the 2′-*O*-ribose-methylated nucleotides and the pseudouridines, is achieved by two distinct families of ribonucleoproteins (RNPs) (reviewed in references 14, 18, 27, 28, and 52). The modification guide RNPs consist of a sequence-specific guide RNA and a set of common proteins. Each 2′-*O*-ribose methylation guide RNA carries the conserved C, C′ (consensus, RUGAUGA), D, and D′ (CUGA) box motifs and possesses one or two 10- to 21-nucleotide-long antisense elements that are responsible for selection of the correct substrate ribonucleotides through the formation of double helices with the target RNAs (11, 32). The selected ribonucleotides are 2′-*O*-ribose methylated by the Nop1p/fibrillarin methyltransferase enzyme that, in addition to the Snu13 (15.5-kDa), Nop56p, and Nop58p RNP proteins, is associated with all box C/D RNAs (14, 18, 27, 52, 55).

The pseudouridylation guide RNAs are composed of two major hairpin elements that are connected by a hinge and followed by a short tail region (Fig. 1A). The single-stranded hinge and tail region carry the conserved H (consensus, ANA NNA) and ACA box motifs that are located at the bases of the 5′ and 3′ hairpins, respectively (6, 20). Two short antisense elements located in an internal loop of the 5′ and/or 3′ hairpins provide the sequence specificity for the pseudouridylation guide RNP by base pairing to the sequences that precede and follow the target uridine. This interaction creates the "pseudouridylation pocket," in which the unpaired substrate uridine selected for pseudouridylation is located 14 or 15 bp upstream of the H or ACA motif of the guide RNA (19, 43). The dyskerin/Cbf5p pseudouridine synthase, together with the Nhp2, Nop10, and Gar1 RNP proteins, is an integral component of box H/ACA pseudouridylation guide RNPs (33, 56).

Vertebrate box H/ACA pseudouridylation guide RNAs are processed from removed and debranched pre-mRNA introns by exonucleolytic activities (18). The conserved H and ACA boxes together with the basal helices of the 5′ and 3′ hairpins provide the signals for correct RNA processing (6, 8, 20). The mature H/ACA RNAs accumulate either in the nucleolus (snoRNAs) or in nucleoplasmic Cajal bodies (small Cajal body-specific RNAs [scaRNAs]). The nucleolar accumulation of box H/ACA snoRNAs is supported by the conserved H and ACA boxes and the basal helix of the 3′ hairpin (34, 42). The box H/ACA scaRNAs carry a common Cajal body-specific localization signal, the CAB box (consensus, UGAG), that is found in the terminal loops of the 5′ and 3′ hairpins (47). In the nucleolus, most box H/ACA snoRNAs direct pseudouridylation of the 18S, 5.8S, and 28S rRNAs (19, 43), while the box H/ACA scaRNAs function in pseudouridylation of the RNA polymerase II (pol II)-transcribed spliceosomal snRNAs (13, 25, 26).

In the yeast *Saccharomyces cerevisiae*, most, if not all, ribosomal pseudouridines are synthesized by box H/ACA snoRNPs

* Corresponding author. Mailing address: Laboratoire de Biologie Moléculaire Eucaryote du CNRS, 118 Rt. de Narbonne, 31062 Toulouse Cedex, France. Phone: 33 5 61 33 59 91. Fax: 33 5 61 33 58 86. E-mail: tamas@ibcg.biotoul.fr.

† Supplemental material for this article may be found at http://mcb.asm.org/.

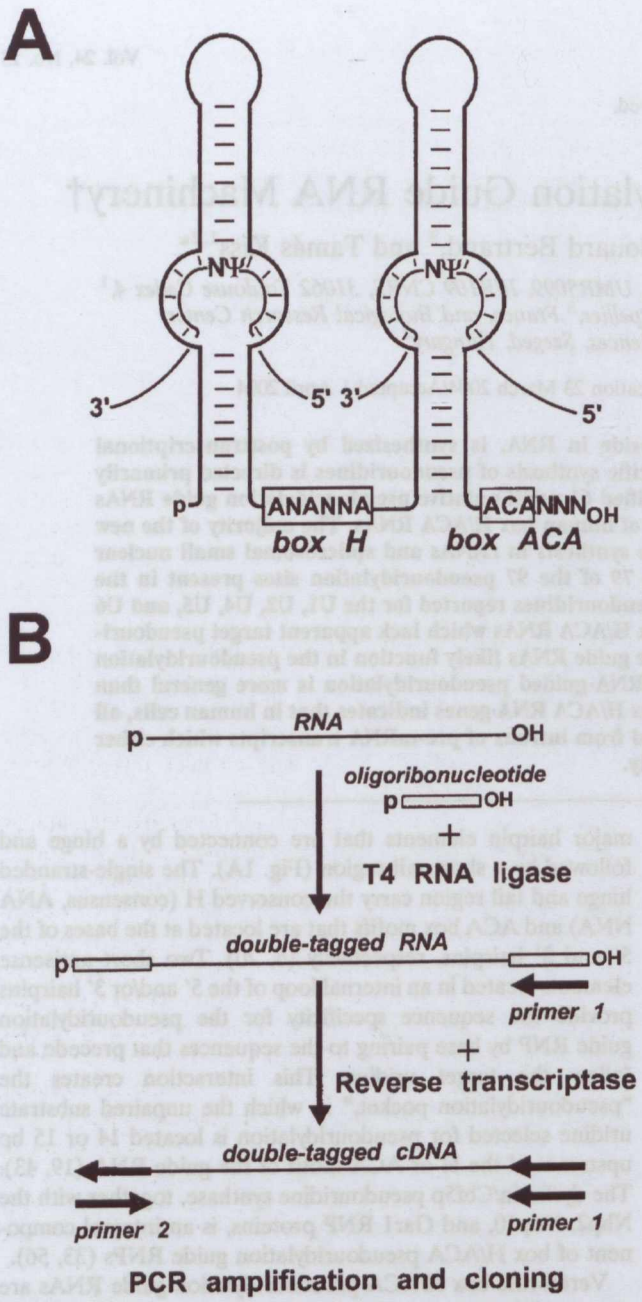


FIG. 1. Schematic structure of box H/ACA RNAs and cDNA construction. (A) Selection of pseudouridylation sites by box H/ACA guide RNAs. For details, see the text. (B) Construction of a cDNA library of human box H/ACA RNAs. HeLa cell RNAs immunoprecipitated by an anti-GAR1 antibody were incubated with a phosphorylated oligoribonucleotide in the presence of T4 RNA ligase. RNA sequences tagged at both termini were converted into double-stranded DNA by a reverse transcription-PCR amplification approach. The amplified DNA was cloned into a plasmid vector, and individual clones were characterized by sequence analysis.

(49). Since only a limited number of box H/ACA RNAs have been identified in humans, it remains unknown to what extent guide RNAs participate in the pseudouridylation of human cellular RNAs (13, 20, 25, 26, 29, 31, 48, 54). A recent identification of partial sequences of several putative box H/ACA snoRNAs in mice suggested that mammalian cells express a large number of box H/ACA RNAs (23). In this study, the identification of 61 novel human box H/ACA RNAs provided us with new insights into the function and organization of the molecular machinery directing the pseudouridylation of human rRNAs, snRNAs, and probably other cellular RNAs.

MATERIALS AND METHODS

Construction and characterization of a cDNA library of human box H/ACA RNAs. Preparation of HeLa cell extracts and immunoprecipitation by a GAR1 antibody of box H/ACA RNPs were performed essentially as described previously (20), except that HeLa cells were sonicated in 40 mM Tris-HCl (pH 7.5) buffer containing 200 mM NaCl and 0.05% Nonidet P-40. The anti-human GAR1 (hGAR1) antipeptide antibody was kindly provided by W. Filipowicz (Friedrich Miescher Institut, Basel, Switzerland). About 0.3 μg of RNA recovered by immunoprecipitation with the anti-hGAR1 antibody was mixed with 40 pmol of 5′-end-phosphorylated oligoribonucleotide (pAAUAAAGCGGCCGC GGAUCCAA) and incubated with 15 U of T4 RNA ligase (Promega) and 10 U of RNase inhibitor (Promega) as described previously (17). After phenol-chloroform extraction, the ligation products were recovered by ethanol precipitation, annealed with 40 pmol of oligodeoxynucleotide P1 (TTGGATCCGCGGCCGC TTTAT), and used as a template for cDNA synthesis with avian myeloblastosis virus reverse transcriptase (Promega). The resulting first-strand cDNA was used as a template for PCR amplification by Vent polymerase (Promega) with oligodeoxynucleotides P1 and P2 (AATAAAGCGGCCGCGGATCCAAA) as primers. The amplified DNA was digested with BamHI, inserted into the BamHI site of pBluescribe (Stratagene), and transformed into *Escherichia coli* DH5α cells. Plasmid purification and sequence analysis were performed according to standard laboratory protocols (50).

Mapping of pseudouridines. Isolation of RNA from human HeLa cells was performed by the guanidine thiocyanate-phenol-chloroform extraction procedure (21). Detection of pseudouridines in the 18S and 28S rRNAs was performed by primer extension analysis of carboxymethyl cellulose (CMC)-alkali-treated HeLa cell RNAs (5). ^{32}P-labeled oligonucleotides complementary to the human 18S rRNA from positions C238 to U256 (Ψ222), U685 to G700 (Ψ613 and Ψ655), C762 to U784 (Ψ690), and A1374 to C1393 (Ψ1330 and Ψ1351) were used as primers. Mapping of Ψ2496 in the 28S rRNA was performed with a primer complementary to the 28S rRNA from positions A2531 to C2548. For numbering of human 18S and 28S rRNAs, see GenBank accession number U13369. The primer extension products were fractionated on 6% sequencing gels.

Expression constructs. The ACA26, ACA35, and ACA57 scaRNAs were overexpressed in human HeLa cells. To this end, the coding regions of ACA26 (oligonucleotides ACTAATCGATTACATTTTGAAGTTAGTGG and TCTA ACGCGTTTGAAATAAGTCAATAAG), ACA35 (oligonucleotides ACTAAT CGATTAGACCTGAGATGTGCTTA and TCTAACGCGTACAGTCACTAA AGCCGTA), and ACA57 (oligonucleotides ACTAATCGATGTAAGTCTGC CTGTCCTAT and TCTAACGCGTCTTAGGACGGCCCTCCTA) were PCR amplified with HeLa cell genomic DNA as a template. The amplified fragments were digested with restriction endonucleases ClaI and XhoI and inserted into the same sites of the pCMV-globin expression construct (13). Transfection of HeLa cells was performed with Fugene 6 (Roche) transfection reagent according to the manufacturer's instructions.

Fluorescence in situ hybridization. Synthesis and chemical conjugation of amino-modified oligodeoxynucleotides with FluoroLink Cy3 monofunctional dye (Amersham), fluorescence hybridization of transfected HeLa cells, and image acquisition and processing were performed as described elsewhere (http://singerlab.aecom.yu.edu) (13). The following oligonucleotide probes were used to detect transiently expressed human scaRNAs (asterisks indicate amino-allyl-modified T residues that are sites of attachment for the fluorescent label): ACA26, AT*C AGCAAAGTCTTACTT*CATCAGACTCAGCCT*T; ACA35, TT*CTTAAA CCCAGCTAT*CACAACACATCACAAGCCTT*T; and ACA57, GT*GTGT CCTGCCAGACT*ACCCTGTTAGAACT*G. A polyclonal rabbit anti-p80-coilin antibody was kindly provided by A. Lamond. Nuclear DNA was stained with 0.1 μg of 4′,6′-diamidino-2-phenylindole/ml.

RESULTS AND DISCUSSION

Identification of novel human box H/ACA RNAs. From a human HeLa cell extract, box H/ACA RNAs were isolated by

immunoprecipitation with an antibody directed against the GAR1 box H/ACA RNP protein (16). Since vertebrate box H/ACA pseudouridylation guide RNAs are processed from pre-mRNA introns (18, 27), the mature RNAs carry a 5′-terminal monophosphosphate and a 3′-terminal hydroxyl group (31). To facilitate the synthesis of full-length cDNAs, the 5′ and 3′ termini of the immunoselected box H/ACA RNAs were extended by the addition of a phosphorylated oligoribonucleotide with the help of T4 RNA ligase (Fig. 1B). Synthesis of cDNA was performed with avian myeloblastosis virus reverse transcriptase and an oligodeoxynucleotide primer complementary to the oligoribonucleotide tag of the template RNA. The resulting doubly-tagged cDNA was PCR amplified, inserted into a plasmid vector, and transformed into *E. coli*. About 1,500 individual clones were analyzed by manual or automated plasmid sequencing.

We identified a total of 1,120 RNA sequences that defined 17 previously identified and 61 novel box H/ACA RNAs. As demonstrated by computer folding (57), the new RNAs displayed all the characteristic hallmarks of box H/ACA RNAs; they folded into a "hairpin-hinge-hairpin-tail" structure and carried H and ACA box motifs (data not shown). Since among the defining structural features of H/ACA RNAs, the presence of an ACA motif or, in a few instances, an AUA motif located 3 nucleotides from the RNA 3′ end was noted, the newly identified RNAs were designated ACA RNAs and numbered from 1 to 61. The expression and size of each RNA were confirmed by Northern blot analysis. The majority of recombinant plasmids (62%) carried cDNAs of full-length box H/ACA RNAs. In a few instances, 5′- and/or 3′-extended RNAs that apparently represented processing intermediates of mature box H/ACA RNAs were also identified. About one-fourth of the recombinant plasmids carried cDNAs corresponding to various fragments of the 18S and 28S rRNAs (15%) or representing full-length or partial sequences of the 5S and 5.8S rRNAs (8%). Less frequently (<2%), we also obtained plasmids carrying spliceosomal snRNA, tRNA, and mRNA sequences. Although about 20 box H/ACA RNAs were highly overrepresented in our cDNA library, we identified the sequences of 18 box H/ACA RNAs only once, suggesting that our survey was not saturated.

The computer-predicted two-dimensional structures of the new box H/ACA RNAs were scrutinized to find putative antisense target recognition elements and eventually to identify potential substrate RNAs. Based on their predicted functions, the new box H/ACA RNAs were divided into three groups. As expected, the majority of these RNAs (43 species) have been implicated in guiding the pseudouridylation of the 18S or 28S rRNAs. Another group of RNAs (6 species) have been assigned to the direction of pseudouridine synthesis for the major spliceosomal snRNAs. Finally, 12 box H/ACA RNAs classified in the third group lacked significant complementarities to any known stable cellular RNAs.

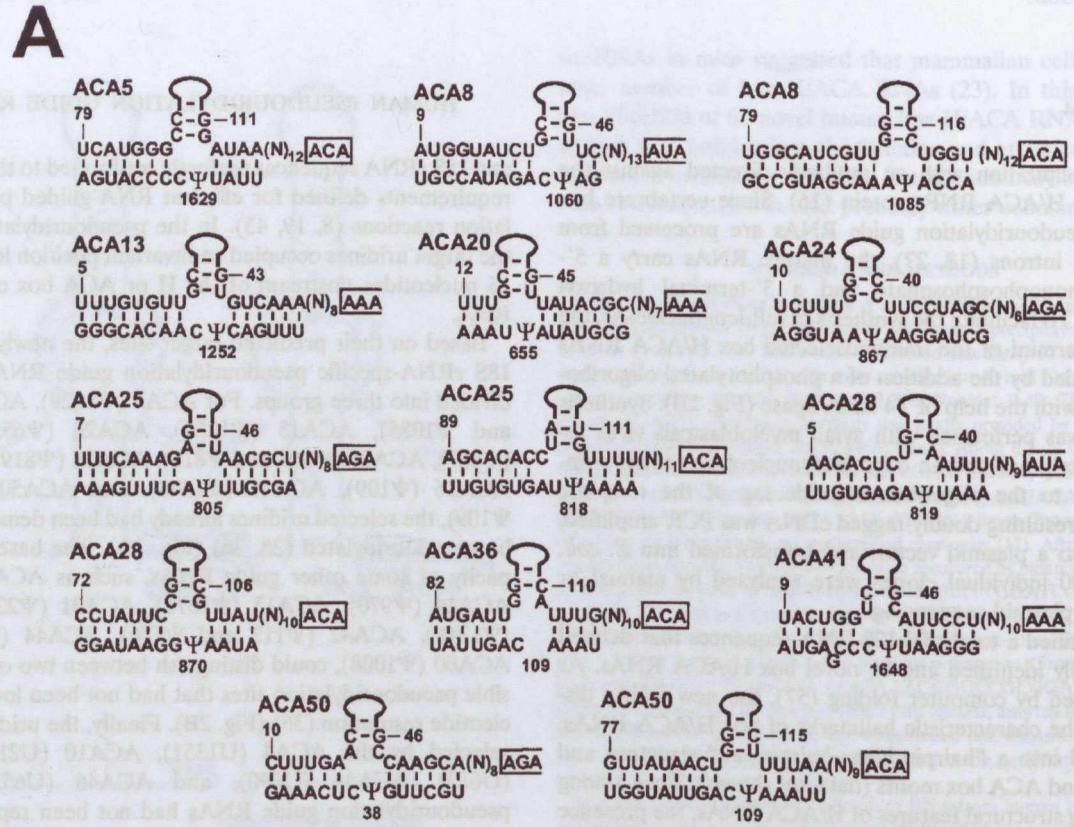
Guide RNAs directing the pseudouridylation of 18S rRNA. The human 18S rRNA is estimated to contain about 38 pseudouridine residues, 30 of which have been located either exactly or to within 2 or 3 nucleotides (35, 36). Of the newly identified box H/ACA RNAs, 19 have been predicted to function in 18S rRNA pseudouridylation (Fig. 2). The potential base-pairing interactions formed between these guide RNAs and 18S rRNA sequences perfectly conformed to the structural requirements defined for efficient RNA-guided pseudouridylation reactions (8, 19, 43). In the pseudouridylation pocket, the target uridines occupied an invariant position located 14 or 15 nucleotides upstream of the H or ACA box of the guide RNA.

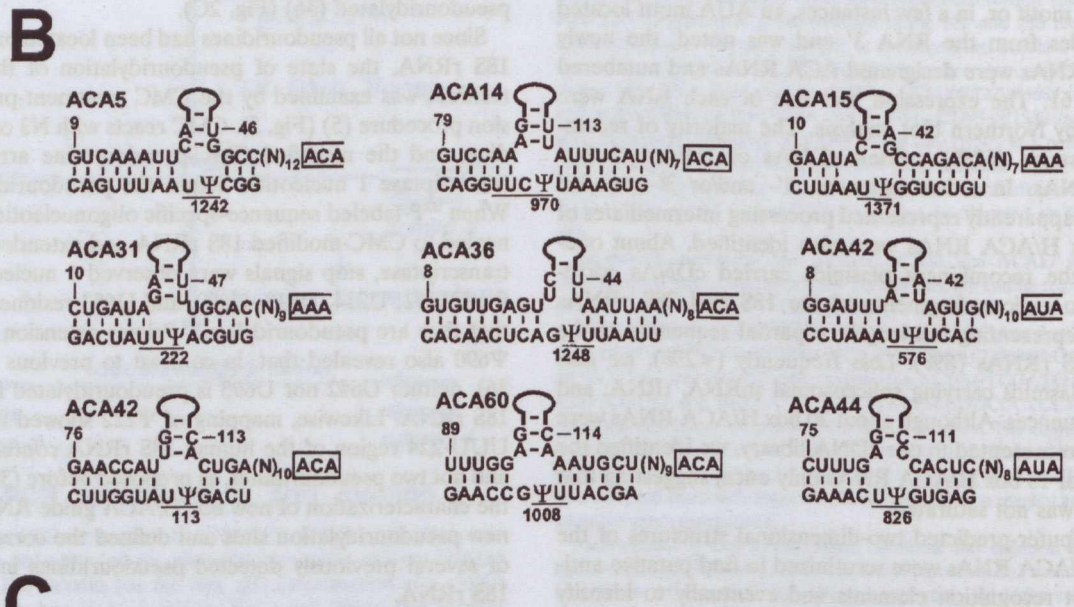
Based on their predicted target sites, the newly discovered 18S rRNA-specific pseudouridylation guide RNAs could be divided into three groups. For ACA5 (Ψ1629), ACA8 (Ψ1060 and Ψ1085), ACA13 (Ψ1252), ACA20 (Ψ655), ACA24 (Ψ867), ACA25 (Ψ805 and Ψ818), ACA28 (Ψ819 and Ψ870), ACA36 (Ψ109), ACA41 (Ψ1648), and ACA50 (Ψ38 and Ψ109), the selected uridines already had been demonstrated to be pseudouridylated (35, 36) (Fig. 2A). The base-pairing capacity of some other guide RNAs, such as ACA5 (Ψ1242), ACA14 (Ψ970), ACA15 (Ψ1371), ACA31 (Ψ222), ACA36 (Ψ1248), ACA42 (Ψ113 and Ψ576), ACA44 (Ψ826), and ACA60 (Ψ1008), could distinguish between two or three possible pseudouridylation sites that had not been located to nucleotide resolution (36) (Fig. 2B). Finally, the uridine residues selected by the ACA4 (U1351), ACA10 (U214), ACA24 (U613), ACA44 (U690), and ACA46 (U653) putative pseudouridylation guide RNAs had not been reported to be pseudouridylated (36) (Fig. 2C).

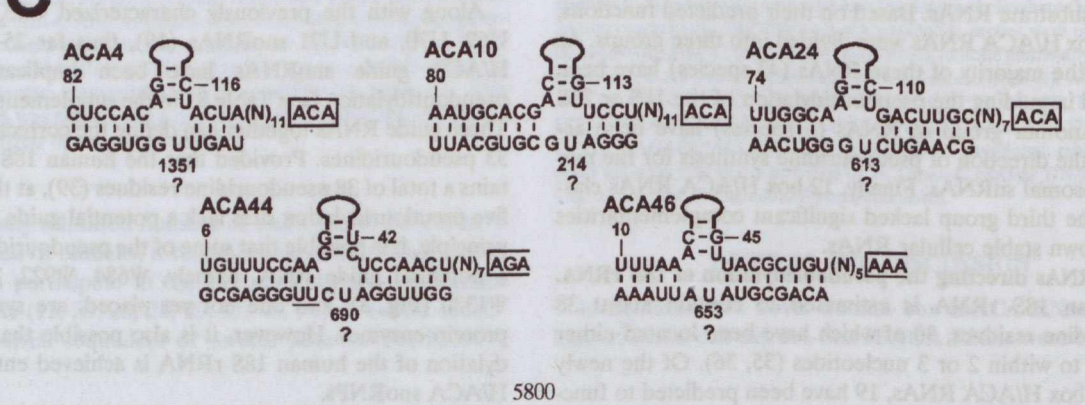
Since not all pseudouridines had been located on the human 18S rRNA, the state of pseudouridylation of these uridine residues was examined by the CMC treatment-primer extension procedure (5) (Fig. 3). CMC reacts with N3 of pseudouridine, and the modified CMC-pseudouridine arrests reverse transcriptase 1 nucleotide before the pseudouridylation site. When ^{32}P-labeled sequence-specific oligonucleotides were annealed to CMC-modified 18S rRNA and extended by reverse transcriptase, stop signals were observed 1 nucleotide before the U1351, U214, U613, U690, and U653 residues, indicating that they are pseudouridylated. Primer extension mapping of Ψ690 also revealed that, in contrast to previous reports (35, 36), neither U692 nor U693 is pseudouridylated in HeLa cell 18S rRNA. Likewise, mapping of Ψ222 showed that the 222-UUU-224 region of the human 18S rRNA contains only one and not two pseudouridines, as proposed before (35, 36). Thus, the characterization of new box H/ACA guide RNAs revealed new pseudouridylation sites and defined the correct positions of several previously detected pseudouridines in the human 18S rRNA.

Along with the previously characterized U23, U66, U67, U69, U70, and U71 snoRNAs (19), thus far 25 human box H/ACA guide snoRNAs have been implicated in 18S pseudouridylation (see Table S1 in the supplemental material). These guide RNAs together can define the correct positions of 33 pseudouridines. Provided that the human 18S rRNA contains a total of 38 pseudouridine residues (39), at this time only five pseudouridylation sites lack a potential guide snoRNA. In principle, it is possible that some of the pseudouridines missing a potential guide RNA, namely, Ψ684, Ψ922, Ψ1178 (36), Ψ1330 (Fig. 3), and one not yet placed, are synthesized by protein enzymes. However, it is also possible that pseudouridylation of the human 18S rRNA is achieved entirely by box H/ACA snoRNPs.

Pseudouridylation guide RNAs directing the modification of 28S rRNA. Based on the estimated ratio of its pseudouridine


A
ACA5
ACA8
ACA8
ACA13
ACA20
ACA24
ACA25
ACA25
ACA28
ACA28
ACA36
ACA41
ACA50
ACA50


B
ACA5
ACA14
ACA15
ACA31
ACA36
ACA42
ACA42
ACA60
ACA44


C
ACA4
ACA10
ACA24
ACA44
ACA46

content to its uridine content, the human 28S rRNA was predicted to carry about 57 pseudouridines (22, 35). Later, primer extension mapping located 54 pseudouridines to nucleotide resolution on the human 28S rRNA (45). At the outset of this study, six human box H/ACA guide snoRNAs (U19, U64, U65 U68, E2, and E3) had been implicated in the synthesis of nine pseudouridines in 28S rRNA (19) (see Table S2 in the supplemental material). Our survey identified 26 additional box H/ACA snoRNAs which could be linked to 35 reported pseudouridylation sites in the 28S rRNA (Fig. 4). Another snoRNA, ACA61, was predicted to position the U2496 residue for pseudouridylation. Indeed, primer extension mapping confirmed the presence of a novel pseudouridine at U2496 (Fig. 3), indicating that ACA61 is a genuine pseudouridylation guide RNA. Therefore, in total, 32 human box H/ACA snoRNAs have been implicated in 28S rRNA pseudouridylation. These guide RNAs can select 44 of the 57 pseudouridylation sites present in the human 28S rRNA (see Table S2 in the supplemental material).

In summary, after identification of 56 putative human rRNA pseudouridylation guide RNAs (19; this study), we can attribute box H/ACA RNA-directed modification to 79 of the estimated 97 pseudouridylation sites present in the human 18S, 5.8S, and 28S rRNAs (see Tables S1 and S2 in the supplemental material). Even allowing the possibility that a few ribosomal pseudouridines are still unknown, we can conclude that the great majority of pseudouridines carried by the human rRNAs are synthesized by box H/ACA RNPs.

Of the 56 guide RNAs implicated in rRNA pseudouridylation, 22 are capable of directing two independent modification reactions. Usually, the two target sites of the "double pseudouridylation guides" are found on the same rRNA and, most frequently, are located close to each other in the primary rRNA sequence. The target pseudouridines selected by the 5′ hairpins of double guides can be located either upstream or downstream of the pseudouridylation sites determined by their 3′ hairpins. Less frequently, the 5′ and 3′ hairpins of a few double guides can function in the pseudouridylation of two rRNA species. The U69 snoRNA can direct the pseudouridylation of the 18S and 5.8S rRNAs, while the ACA10 and ACA31 snoRNAs are predicted to function in the modification of the 18S and 28S rRNAs.

Together, the two short helixes formed by the H/ACA guide RNA and rRNA sequences preceding and following the target uridine comprise a minimum of 9 bp or a maximum of 16 bp (Fig. 2 and 4). Most frequently (in 31% of the total instances), the rRNA-H/ACA RNA interaction involves 11 bp. In a few instances, mismatched (ACA24, ACA25, ACA15, and ACA58) or bulging (ACA41) nucleotides are also involved in the predicted rRNA-H/ACA RNA interaction. It is also noteworthy that G-U base pairs frequently occur in the predicted rRNA-H/ACA RNA helices, especially those composed of more than 10 bp.

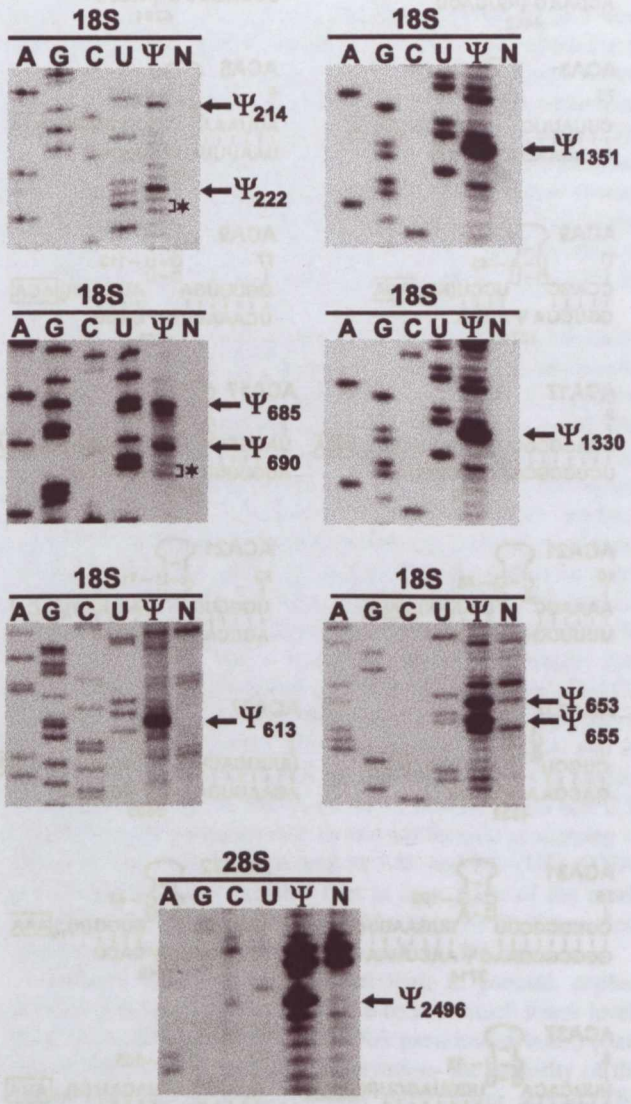


FIG. 3. Verification of pseudouridine residues in human 18S and 28S rRNAs predicted by guide RNA-rRNA interactions. CMC-alkali-modified (Ψ) or control (N) HeLa cell RNAs were analyzed by primer extension with ^{32}P-labeled oligonucleotide primers complementary to the appropriate regions of the human 18S and 28S rRNAs. Lanes A, G, C, and U show dideoxy sequencing reactions performed on recombinant plasmids carrying the human 18S or 28S rRNA genes. Brackets and asterisks indicate uridines that were reported to be pseudouridylated.

Identification of the ACA2 and ACA34 snoRNAs provided

FIG. 2. Potential base-pairing interactions between box H/ACA RNAs and human 18S rRNA. (A) Selection of known pseudouridylation sites. The upper strands represent box H/ACA RNA sequences in a 5′-to-3′ orientation. Solid lines represent the upper parts of the 5′ or 3′ hairpins of guide RNAs. The ACA motifs are in closed boxes. The first three nucleotides of the putative H motifs are in open-ended boxes. The lower strands represent 18S rRNA sequences in a 3′-to-5′ orientation. The positions of pseudouridine residues were reported previously (36). The pseudouridine residues defined by interactions with guide RNAs are indicated (Ψ). The sequence of human 18S rRNA is from GenBank accession number U13369. (B) H/ACA RNAs distinguishing between two or three potential pseudouridylation sites. A bar below rRNA sequences indicates that one of the underlined uridines is pseudouridine. (C) Prediction of new pseudouridylation sites. Question marks indicate novel pseudouridylation sites revealed by identification of the corresponding guide RNAs.

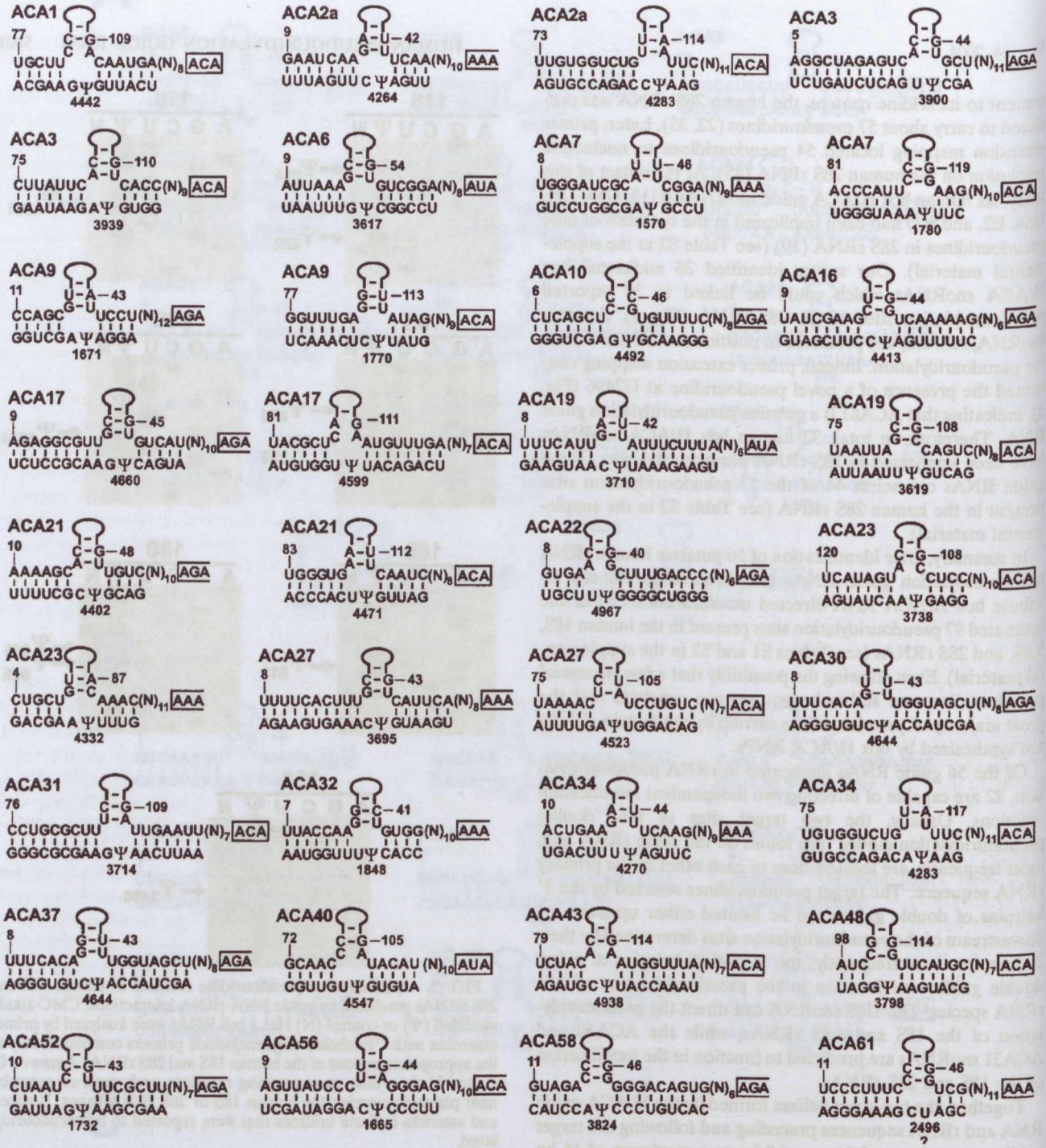


FIG. 4. Potential base-pairing interactions between box H/ACA RNAs and human 28S rRNA. The sequence of human 28S rRNA is from GenBank accession number U13369. The positions of pseudouridine residues were reported previously (45). See the legend to Fig. 2 for other details.

us with new insights into the molecular mechanism of the evolution of modification guide RNAs. The ACA34 snoRNA and two sequence variants of the ACA2 snoRNA (ACA2a and ACA2b) are encoded within three different introns of a hypothetical protein gene, FLJ20436 (see Table S2 in the supplemental material). Interestingly, the ACA34 snoRNA shows a strong sequence similarity to both isoforms of ACA2 (66% identity). Therefore, it could be considered a third sequence variant of ACA2. Consistent with this notion, the 3′ hairpin of ACA34, similar to those of ACA2a and ACA2b, can position the U4283 residue in the 28S rRNA for pseudouridylation (Fig. 4). However, while the 5′ hairpins of ACA2a and ACA2b

can direct the synthesis of Ψ4264, the 5′ hairpin of ACA34 selects Ψ4270 in the 28S rRNA. Apparently, the ACA2a, ACA2b, and ACA34 snoRNA genes have been generated by subsequent gene duplications during evolution. After the first duplication event, random point mutations in the target recognition motifs of the ACA34 gene or the parental gene of the contemporary ACA2a and ACA2b genes resulted in a novel snoRNA gene with new sequence specificity. The second gene duplication event generated the current ACA2a and ACA2b genes. Of course, random mutations in the ACA2a or ACA2b genes may produce another, functionally distinct pseudouridylation guide RNA gene during future evolution.

Guide RNAs implicated in the pseudouridylation of spliceosomal snRNAs. The human U1, U2, U4, U5, and U6 spliceosomal snRNAs together carry 21 pseudouridines (38). Earlier, we characterized four guide RNAs (U85, U89, U92, and U93) which were predicted to direct the synthesis of four pseudouridines in the U5 (Ψ85 and Ψ89) and U2 (Ψ92 and Ψ93) snRNAs (see Table S3A in the supplemental material). The current survey identified six additional box H/ACA RNAs implicated in the pseudouridylation of the U6 (ACA12), U2 (ACA26, ACA35, and ACA45), U1 (ACA47), and U5 (ACA57) snRNAs (Fig. 5A). Together, the above-described H/ACA guide RNAs can direct the synthesis of 11 pseudouridine residues in the U1, U2, U5, and U6 spliceosomal snRNAs. This finding indicates that the involvement of box H/ACA guide RNAs in snRNA pseudouridylation is more general than demonstrated before. However, at this time, we cannot exclude the possibility that protein enzymes also contribute to the pseudouridylation of spliceosomal snRNAs (25, 37).

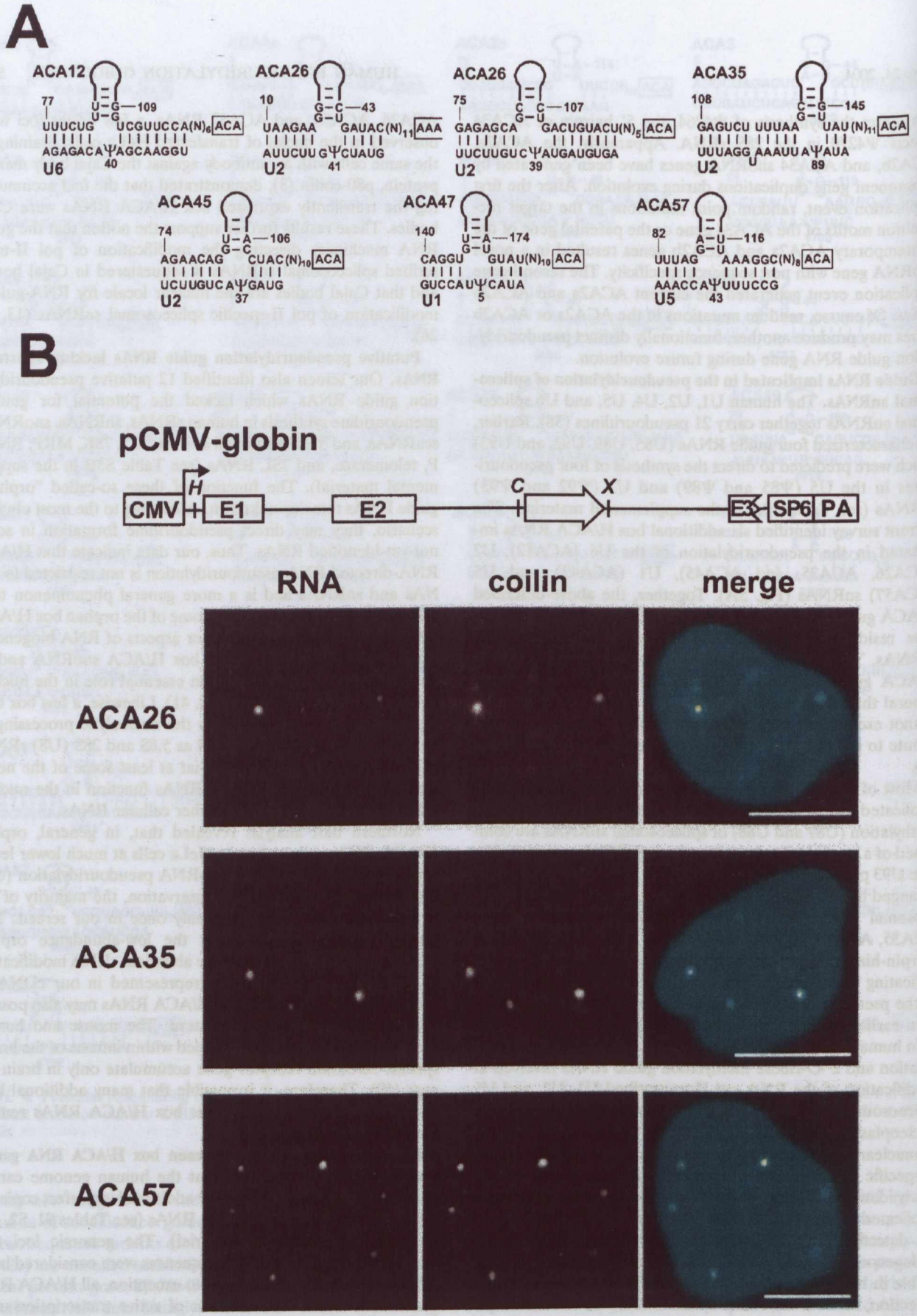
Most of the previously identified modification guide RNAs implicated in pseudouridylation (U85 and U89) or 2′-*O*-ribose methylation (U87 and U88) of spliceosomal snRNAs are composed of a box H/ACA domain and a box C/D domain (13, 25). The U93 pseudouridylation guide RNA contains two tandemly arranged box H/ACA domains (26). In contrast, the new spliceosomal pseudoridylation guide RNAs, ACA12, ACA26, ACA35, ACA45, ACA47, and ACA57, possess the consensus hairpin-hinge-hairpin-tail structure of box H/ACA snoRNAs, indicating that canonical box H/ACA guide RNAs participate in the pseudouridylation of spliceosomal snRNAs more often than earlier observations suggested (13).

In human HeLa cells, the formerly characterized pseudouridylation and 2′-*O*-ribose methylation guide RNAs involved in modification of the RNA pol II-transcribed U1, U2, and U5 spliceosomal snRNAs were found to specifically accumulate in nucleoplasmic Cajal bodies (13, 24, 26). To further explore the subnuclear organization of the modification machinery of pol II-specific snRNAs, we investigated the localization of the newly identified ACA26, ACA35, and ACA57 RNAs that were implicated in the pseudouridylation of the U2 snRNA. Due to the detection limit of fluorescence in situ hydridization, the endogenous ACA26, ACA35, and ACA57 RNAs were not visible in HeLa cells (data not shown). Therefore, to facilitate detection, the ACA26, ACA35, and ACA57 RNAs were transiently overexpressed in HeLa cells by using the pCMV-globin expression construct, which had been developed to study the expression of intronic snoRNAs (13, 30) (Fig. 5B). Upon hybridization with fluorescent oligonucleotides specific for the ACA26, ACA35, and ACA57 RNAs, a few bright foci were observed in the nuclei of transfected HeLa cells. Staining of the same cells with an antibody against the Cajal body marker protein, p80-coilin (3), demonstrated that the foci accumulating the transiently expressed box H/ACA RNAs were Cajal bodies. These results further support the notion that the guide RNA machinery directing the modification of pol II-transcribed spliceosomal snRNAs is sequestered in Cajal bodies and that Cajal bodies are the nuclear locale for RNA-guided modification of pol II-specific spliceosomal snRNAs (13, 24, 26).

Putative pseudouridylation guide RNAs lacking substrate RNAs. Our screen also identified 12 putative pseudouridylation guide RNAs which lacked the potential for guiding pseudouridine synthesis in human rRNAs, snRNAs, snoRNAs, scaRNAs, and YRNAs as well as in the U7, 7SK, MRP, RNase P, telomerase, and 7SL RNAs (see Table S3B in the supplemental material). The function of these so-called "orphan" guide RNAs remains unknown. According to the most obvious scenario, they may direct pseudouridine formation in some not-yet-identified RNAs. Thus, our data indicate that H/ACA RNA-directed RNA pseudouridylation is not restricted to rRNAs and snRNAs and is a more general phenomenon than assumed before. Alternatively, some of the orphan box H/ACA RNAs may also function in other aspects of RNA biogenesis. For example, the human U17 box H/ACA snoRNA and its yeast orthologue, snR30, play an essential role in the nucleolytic processing of 18S rRNA (4, 41). Likewise, a few box C/D snoRNAs play a crucial role in the nucleolytic processing of 18S (U3, U14, and U22) as well as 5.8S and 28S (U8) rRNAs (52). Therefore, it is possible that at least some of the newly identified orphan box H/ACA RNAs function in the nucleolytic processing of rRNAs or other cellular RNAs.

Northern blot analysis revealed that, in general, orphan H/ACA RNAs accumulate in HeLa cells at much lower levels than do snoRNAs involved in rRNA pseudouridylation (data not shown). In line with this observation, the majority of the new orphan RNAs appeared only once in our screen. This finding strongly suggests that the low-abundance orphan H/ACA RNAs, in contrast to the abundant rRNA modification guide RNAs, are highly underrepresented in our cDNA library. On the other hand, box H/ACA RNAs may also possess a tissue-specific expression pattern. The mouse and human HBI-36 snoRNAs that are encoded within introns of the brain-specific serotonin receptor gene accumulate only in brain tissues (10). Therefore, it is possible that many additional low-abundance and/or tissue-specific box H/ACA RNAs remain unidentified in human cells.

Genomic organization of human box H/ACA RNA genes. Database searches revealed that the human genome carries one perfect and sometimes a few additional imperfect copies of the newly identified box H/ACA RNAs (see Tables S1, S2, and S3 in the supplemental material). The genomic loci that matched perfectly our cDNA sequences were considered bona fide H/ACA RNA genes. With no exception, all H/ACA RNA genes were found within introns of active transcription units known to produce spliced mRNAs. Moreover, all bona fide H/ACA RNA genes showed a parallel orientation with their host genes, indicating that the H/ACA RNAs and their host pre-mRNAs are synthesized cotranscriptionally. The majority


A
ACA12
77
109
UUUCUG
UCGUUCCA(N)6 ACA
AGAGACAΨAGCAAGGU
40
U6
ACA26
10
43
UAAGAA
GAUAC(N)11 AAA
AUUCUU GΨCUAUG
41
U2
ACA26
75
107
GAGAGCA
GACUGUACU(N)5 ACA
UUCUUGUCΨAUGAUGUGA
39
U2
ACA35
108
145
GAGUCU UUUAA
UAU(N)11 ACA
UUUAGG AAAUU AΨAUA
U
89
U2
ACA45
74
106
AGAACAG
CUAC(N)10 ACA
UCUUGUCAΨGAUG
37
U2
ACA47
140
174
CAGGU
GUAU(N)10 ACA
GUCCAUΨCAUA
5
U1
ACA57
82
116
UUUAG
AAAGGC(N)8 ACA
AAACCAΨUUUCCG
43
U5
B
pCMV-globin
H
CMV
E1
E2
C
X
E3
SP6
PA
RNA
coilin
merge
ACA26
ACA35
ACA57

of the H/ACA host genes are protein-coding genes, although in many instances, the functions of their predicted protein products remain unknown. Frequently, two or even more H/ACA RNAs are encoded within different introns of the same host gene. For example, the MGC5306 gene, which encodes a hypothetical protein, hosts at least six box H/ACA RNAs (ACA1, ACA8, ACA18, ACA25, ACA32, and ACA40) as well as two box C/D snoRNAs (mgh28S-2410 and mgh28S-2412) that are predicted to direct 2′-*O*-ribose methylation of the human 28S rRNA at positions C2410 and G2412 (unpublished results).

The rRNA pseudouridylation guide RNA genes are frequently located in ribosomal protein genes or in other genes connected to ribosome biogenesis (nucleolin) or protein synthesis (translation initiation and elongation factors and cytoplamic protein chaperones). Interestingly, the ACA36 and ACA56 genes lie within introns of the dyskerin gene, which encodes the common pseudouridine synthase of box H/ACA RNPs. Another snoRNA, ACA23, is hosted by the importin 7 gene, which encodes an import receptor for ribosomal proteins. Collectively, these observations further corroborate the notion that cotranscription is an important way of coordinating the regulation of factors required for ribosome synthesis and function. Therefore, it is possible that at least some of the snoRNA host genes that lack a function will be shown to participate in some aspects of protein synthesis. On the other hand, there are also a few host genes, for example, the cytochrome P450 oxidoreductase (ACA14a and ACA14b), methyl-CpG binding domain protein 2 (a DNA methyltransferase) (ACA37), and SRCAP (a transcriptional activator) (ACA30) genes, that cannot be directly linked to ribosome biogenesis or function.

Usually, the host genes of H/ACA RNAs directing the pseudouridylation of spliceosomal snRNAs (see Table S3A in the supplemental material) or lacking potential substrate RNAs (see Table S3B in the supplemental material) cannot be directly connected to ribosome biogenesis or translation. The ACA33 and ACA51 RNAs represent the only exceptions to this rule, since they are encoded within the S12 ribosomal protein and NOP56 box C/D snoRNP protein genes, respectively. The nonribosomal guide RNA genes frequently appear within genes encoding proteins involved in the functional maintenance of the human genome, such as the catalytic subunit of DNA polymerase alpha (ACA12), condensin subunit 1 (U85), nucleosome assembly protein 1-like protein (ACA54), and chromodomain helicase DNA binding protein 4 (ACA57). The functional significance of this observation is still unclear.

Several box H/ACA RNAs are encoded within genes that have little capacity for protein coding. The spliced and polyadenylated mRNA-like products of these genes contain no long open reading frames. Therefore, it appears that the only function of these genes is to express their intronic H/ACA RNAs (53). Of the 75 known human box H/ACA RNA genes, 15 seem to be located within introns of non-protein-coding genes, indicating that cotranscription within non-protein-coding pre-mRNAs is a rather common way to express pseudouridylation guide RNAs. Previously, four non-protein-coding snoRNA host genes were identified and shown to belong to the family of 5′-terminal oligopyrimidine (5′TOP) genes (9, 46, 51, 53). The sequence of 5′TOP mRNAs commences with a 5′-terminal C residue and is followed by a short pyrimidine tract that plays an important role in the upregulation of transcription and translation of 5′TOP mRNAs (2). The family of 5′TOP genes also includes ribosomal protein and translation elongation factor genes. The expression of rRNA modification guide snoRNAs within non-protein-coding 5′TOP pre-mRNAs therefore may provide a regulatory mechanism to coordinate the accumulation of snoRNAs and ribosomal proteins. Inspection of the 5′-terminal sequences of the expressed sequence tags of the newly identified non-protein-coding H/ACA snoRNA host genes revealed that at least two of them, named TOP1 (encoding ACA16, ACA44, and ACA61) and TOP2 (encoding ACA17 and ACA43), belong to the family of 5′TOP genes. Unfortunately, the correct 5′ ends of other non-protein-coding H/ACA host genes could not be inferred from their partial expressed sequence tags deposited in databases.

Besides the bona fide genes, most box H/ACA RNAs possess one or more imperfect genomic copies. These defective copies often represent 5′- or 3′-truncated versions of the authentic H/ACA RNA gene, indicating that they are apparently pseudogenes. Consistent with this conclusion, the sequences of the truncated H/ACA genes cannot be folded into the characteristic secondary structures of box H/ACA RNAs. Other genomic copies represent full-length H/ACA RNAs, but they contain numerous point mutations, short internal deletions, and/or insertions. These genomic sequences frequently fail to fold into a perfect hairpin-hinge-hairpin-tail structure or lack functional H and/or ACA motifs, suggesting that they do not code for functional RNAs. Supporting this notion, the default H/ACA sequences are frequently located in transcriptionally silent genomic loci or, alternatively, lie within known transcription units but in an opposite orientation. Both truncated and full-length pseudogene sequences are often followed by 8- to 15-nucleotide-long oligo(A) tracts and sometimes are flanked by short perfect repeats, indicating that they have been generated by retrotransposition (54). Finally, full-length H/ACA RNA sequences carrying a few point mutations are also found within introns of known genes in a parallel orientation. Since the sequences of these genes fold into the hairpin-hinge-hairpin-tail structure, they likely represent functional H/ACA

FIG. 5. Putative pseudouridylation guide RNAs directing the modification of spliceosomal snRNAs. (A) Proposed base-pairing interactions between box H/ACA guide RNAs and human spliceosomal snRNAs. (B) Fluorescence in situ localization of guide RNAs transiently overexpressed in HeLa cells. The schematic structure of the pCMV-globin expression construct is shown. The promoter region of cytomegalovirus (CMV), the exons in the human β-globin gene (E1 to E3), the polyadenylation region in the bovine growth hormone gene (PA), and the SP6 promoter are shown. The relevant restriction sites are indicated (*H*, HindIII; *C*, ClaI; *X*, XhoI). Fluorescence in situ hybridization with oligonucleotide probes specific for the ACA26, ACA35, and ACA57 scaRNAs was combined with indirect immunofluorescence with an antibody against the Cajal body marker protein, p80-coilin. The nuclear DNA was stained with 4′,6′-diamidino-2-phenylindole (blue). Bar, 10 μm.

RNA genes expressing sequence variants of the characterized RNAs.

Conclusions. In this study, we identified 61 human box H/ACA snoRNAs and scaRNAs. The majority of these RNAs (49 species) are predicted to function as guide RNAs in the synthesis of 71 pseudouridine residues in the human 18S and 28S rRNAs and the U1, U2, U5, and U6 spliceosomal snRNAs. Some of the new H/ACA RNAs (12 species) lack potential target sites in human rRNAs, snRNAs, and snoRNAs. These orphan H/ACA RNAs either direct the pseudouridylation of some not-yet-identified RNAs or function in other aspects of cellular RNA biogenesis. As predicted by their genomic organization, all human box H/ACA RNAs are processed from pre-mRNA introns. The host genes of human H/ACA RNAs can be divided into three major groups. Most of them encode well-characterized proteins that frequently function in ribosome biogenesis or protein synthesis. Another group of H/ACA RNA host genes are predicted to encode proteins with unknown functions. Finally, the host genes of many H/ACA RNAs lack apparent protein-coding capacity and frequently belong to the family of 5′TOP genes.

ACKNOWLEDGMENTS

We are grateful to Y. de Préval for the synthesis of oligodeoxynucleotides. We thank A. I. Lamond and W. Filipowicz for providing us with antibodies against p80-coilin and hGAR1. We thank M. Weber for critical reading of the manuscript and also for constructing a GenBank custom track of human H/ACA RNAs.

A.M.K. was funded by a short-term EMBO fellowship, a Hungarian State Eötvös fellowship, and l'Association pour la Recherche contre le Cancer. This work was supported by grants from CNRS, la Ligue Nationale contre le Cancer, and the Hungarian Research Foundation (OTKA, T31738).

REFERENCES

1. **Agris, P. F.** 1996. The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function. Prog. Nucleic Acids Res. Mol. Biol. **53**:79–129.
2. **Amaldi, F., and P. Pierandrei-Amaldi.** 1997. TOP genes: a translationally controlled class of genes including those coding for ribosomal proteins. Prog. Mol. Subcell. Biol. **18**:1–17.
3. **Andrade, L. E., E. K. Chan, I. Raska, C. L. Peebles, G. Roos, and E. M. Tan.** 1991. Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. J. Exp. Med. **173**:1407–1419.
4. **Atzorn, V., P. Fragapane, and T. Kiss.** 2003. U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. Mol. Cell. Biol. **24**:1769–1778.
5. **Bakin, A., and J. Ofengand.** 1993. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. Biochemistry **32**:9754–9762.
6. **Balakin, A. G., L. Smith, and M. J. Fournier.** 1996. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. Cell **86**:823–834.
7. **Björk, G. R., J. U. Ericson, C. E. D. Gustafsson, T. G. Hagervall, Y. H. Jönsson, and P. M. Wikström.** 1987. Transfer RNA modification. Annu. Rev. Biochem. **56**:263–287.
8. **Bortolin, M. L., P. Ganot, and T. Kiss.** 1999. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. EMBO J. **18**:457–469.
9. **Bortolin, M. L., and T. Kiss.** 1998. Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. RNA **4**:445–454.
10. **Cavaillé, J., K. Buiting, M. Kiefmann, M. Lalande, C. I. Brannan, B. Horsthemke, J. P. Bachellerie, J. Brosius, and A. Hüttenhofer.** 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. USA **97**:14311–14316.
11. **Cavaillé, J., M. Nicoloso, and J. P. Bachellerie.** 1996. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. Nature **383**:732–735.
12. **Charette, M., and M. W. Gray.** 2000. Pseudouridine in RNA: what, where, how, and why. IUBMB Life **49**:341–351.
13. **Darzacq, X., B. E. Jády, C. Verheggen, A. M. Kiss, E. Bertrand, and T. Kiss.** 2002. Cajal body-specific small nuclear RNAs: a novel class of 2′-O-methylation and pseudouridylation guide RNAs. EMBO J. **21**:2746–2756.
14. **Decatur, W. A., and M. J. Fournier.** 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. J. Biol. Chem. **278**:695–698.
15. **Decatur, W. A., and M. J. Fournier.** 2002. rRNA modifications and ribosome function. Trends Biochem. Sci. **27**:344–351.
16. **Dragon, F., V. Pogacic, and W. Filipowicz.** 2000. In vitro assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. Mol. Cell. Biol. **20**:3037–3048.
17. **England, T., A. Bruce, and O. Uhlenbeck.** 1980. Specific labeling of 3′ termini of RNA with T4 RNA ligase. Methods Enzymol. **65**:65–74.
18. **Filipowicz, W., and V. Pogacic.** 2002. Biogenesis of small nucleolar ribonucleoproteins. Curr. Opin. Cell Biol. **14**:319–327.
19. **Ganot, P., M. L. Bortolin, and T. Kiss.** 1997. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell **89**:799–809.
20. **Ganot, P., M. Caizergues-Ferrer, and T. Kiss.** 1997. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. Genes Dev. **11**:941–956.
21. **Goodall, G. J., K. Wiebauer, and W. Filipowicz.** 1990. Analysis of pre-mRNA processing in transfected plant protoplasts. Methods Enzymol. **181**:148–161.
22. **Hughes, D. G., and B. E. Maden.** 1978. The pseudouridine contents of the ribosomal ribonucleic acids of three vertebrate species. Numerical correspondence between pseudouridine residues and 2′-*O*-methyl groups is not always conserved. Biochem. J. **171**:781–786.
23. **Hüttenhofer, A., M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach, J. P. Bachellerie, and J. Brosius.** 2001. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J. **20**:2943–2953.
24. **Jády, B. E., X. Darzacq, K. E. Tucker, A. G. Matera, E. Bertrand, and T. Kiss.** 2003. Modification of small nuclear RNAs occurs in the nucleoplasmic Cajal bodies following import from the cytoplasm. EMBO J. **22**:1878–1888.
25. **Jády, B. E., and T. Kiss.** 2001. A small nucleolar guide RNA functions both in 2′-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. EMBO J. **20**:541–551.
26. **Kiss, A. M., B. E. Jády, X. Darzacq, C. Verheggen, E. Bertrand, and T. Kiss.** 2002. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. Nucleic Acids Res. **30**:4643–4649.
27. **Kiss, T.** 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J. **20**:3617–3622.
28. **Kiss, T.** 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell **109**:145–148.
29. **Kiss, T., M. L. Bortolin, and W. Filipowicz.** 1996. Characterization of the intron-encoded U19 RNA, a new mammalian small nucleolar RNA that is not associated with fibrillarin. Mol. Cell. Biol. **16**:1391–1400.
30. **Kiss, T., and W. Filipowicz.** 1995. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. Genes Dev. **9**:1411–1424.
31. **Kiss, T., and W. Filipowicz.** 1993. Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. EMBO J. **12**:2913–2920.
32. **Kiss-László, Z., Y. Henry, J. P. Bachellerie, M. Caizergues-Ferrer, and T. Kiss.** 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell **85**:1077–1088.
33. **Lafontaine, D. L., C. Bousquet-Antonelli, Y. Henry, M. Caizergues-Ferrer, and D. Tollervey.** 1998. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. **12**:527–537.
34. **Lange, T. S., M. Ezrokhi, F. Amaldi, and S. A. Gerbi.** 1999. Box H and box ACA are nucleolar localization elements of U17 small nucleolar RNA. Mol. Biol. Cell **10**:3877–3890.
35. **Maden, B. E.** 1990. The numerous modified nucleotides in eukaryotic ribosomal RNA. Prog. Nucleic Acids Res. Mol. Biol. **39**:241–303.
36. **Maden, E. H., and J. A. Wakeman.** 1988. Pseudouridine distribution in mammalian 18 S ribosomal RNA. A major cluster in the central region of the molecule. Biochem. J. **249**:459–464.
37. **Massenet, S., Y. Motorin, D. L. Lafontaine, E. C. Hurt, H. Grosjean, and C. Branlant.** 1999. Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase Pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. Mol. Cell. Biol. **19**:2142–2154.
38. **Massenet, S., A. Mougin, and C. Branlant.** 1998. Posttrancriptional modification in the U small nuclear RNAs, p. 201–227. *In* H. Grosjean and R. Benne (ed.), Modification and editing of RNA. ASM Press, Washington, D.C.
39. **McCallum, F. S., and B. E. Maden.** 1985. Human 18 S ribosomal RNA sequence inferred from DNA sequence. Variations in 18 S sequences and secondary modification patterns between vertebrates. Biochem. J. **232**:725–733.
40. **McCloskey, J. A., and P. F. Crain.** 1998. The RNA modification database—1998. Nucleic Acids Res. **26**:196–197.

41. **Morrissey, J. P., and D. Tollervey.** 1993. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. Mol. Cell. Biol. **13**:2469–2477.
42. **Narayanan, A., A. Lukowiak, B. E. Jády, F. Dragon, T. Kiss, R. M. Terns, and M. P. Terns.** 1999. Nucleolar localization signals of box H/ACA small nucleolar RNAs. EMBO J. **18**:5120–5130.
43. **Ni, J., A. L. Tien, and M. J. Fournier.** 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell **89**:565–573.
44. **Ofengand, J.** 2002. Ribosomal RNA pseudouridines and pseudouridine synthases. FEBS Lett. **514**:17–25.
45. **Ofengand, J., and A. Bakin.** 1997. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. J. Mol. Biol. **266**:246–268.
46. **Pelczar, P., and W. Filipowicz.** 1998. The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5′-terminal oligopyrimidine gene family. Mol. Cell. Biol. **18**:4509–4518.
47. **Richard, P., X. Darzacq, E. Bertrand, B. E. Jády, C. Verheggen, and T. Kiss.** 2003. A common sequence motif determines the Cajal body-specific localisation of box H/ACA scaRNAs. EMBO J. **22**:4283–4293.
48. **Ruff, E. A., O. J. Rimoldi, B. Raghu, and G. L. Eliceiri.** 1993. Three small nucleolar RNAs of unique nucleotide sequences. Proc. Natl. Acad. Sci. USA **90**:635–638.
49. **Samarsky, D. A., and M. J. Fournier.** 1999. A comprehensive database for the small nucleolar RNAs from Saccharomyces cerevisiae. Nucleic Acids Res. **27**:161–164.
50. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
51. **Smith, C. M., and J. A. Steitz.** 1998. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. Mol. Cell. Biol. **18**:6897–6909.
52. **Terns, M. P., and R. M. Terns.** 2002. Small nucleolar RNAs: versatile *trans*-acting molecules of ancient evolutionary origin. Gene Expr. **10**:17–39.
53. **Tycowski, K. T., M. D. Shu, and J. A. Steitz.** 1996. A mammalian gene with introns instead of exons generating stable RNA products. Nature **379**:464–466.
54. **Vitali, P., H. Royo, H. Seitz, J. P. Bachellerie, A. Hüttenhofer, and J. Cavaillé.** 2003. Identification of 13 novel human modification guide RNAs. Nucleic Acids Res. **31**:6543–6551.
55. **Wang, H., D. Boisvert, K. K. Kim, R. Kim, and S. H. Kim.** 2000. Crystal structure of a fibrillarin homologue from Methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution. EMBO J. **19**:317–323.
56. **Zebarjadian, Y., T. King, M. J. Fournier, L. Clarke, and J. Carbon.** 1999. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. Mol. Cell. Biol. **19**:7461–7472.
57. **Zuker, M., D. H. Mathews, and D. H. Turner.** 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide, p. 11–43. *In* J. Barciszewski and B. F. C. Clark (ed.), RNA biochemistry and biotechnology. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Table 1

RNA	Length	Target	Host gene	Additional copies	Ref.	Accession No
		18S rRNA				
U23	138	Ψ97	nucleolin	4	(20)	AJ007015
U66	133	Ψ123	ribosomal protein L5	-	(20)	Y11160
U67	137	Ψ1449	eukaryotic translation initiation factor 4A1	3	(20)	Y11161
U69*	132	Ψ40	ribosomal protein L39	1	(20)	Y11163
U70	135	Ψ1698	ribosomal protein L10	15	(20)	Y11164
U71a/b	139	Ψ410	spliced ESTs (no apparent protein)	1	(20)	Y11165/6
ACA4	138	Ψ1351	eukaryotic translation initiation factor 4A2	-	this study	AJ609452
ACA5a/b/c	134	Ψ1242, Ψ1629	hypothetical protein KIAA0948	-	this study	AJ609453
ACA8	139	Ψ1060, Ψ1085	hypothetical protein MGC5306	3	this study	AJ609454
ACA10*	133	Ψ214	ribosomal protein S2	1	this study	AJ609432
ACA13	133	Ψ1252	hypothetical protein TIGA1	-	this study	AJ609455
ACA14a/b	134/135	Ψ970	P450 (cytochrome) oxidoreductase/TOMM20-PENDING	1	this study	AJ609456/7
ACA15	133	Ψ1371	TCP1, zeta subunit (cytopl. protein chaperone)	3	this study	AJ609458
ACA20	132	Ψ655	TCP1, alpha subunit (cytopl. protein chaperone)	2	this study	AJ609459
ACA24	131	Ψ867, Ψ613	spliced ESTs (no apparent protein)	2	this study	AJ609460
ACA25	134	Ψ805, Ψ818	hypothetical protein MGC5306	5	this study	AJ609461
ACA28	127	Ψ819, Ψ870	eukaryotic translation initiation factor 5	-	this study	AJ609462
ACA31*	130	Ψ222	hypothetical protein TPT1	1	this study	AJ609441
ACA36	132	Ψ1248, Ψ109	dyskerin	3	this study	AJ609463
ACA41	132	Ψ1648	eukaryotic translation elongation factor 1-beta	1	this study	AJ609464
ACA42	134	Ψ576, Ψ113	hypothetical protein KIAA0907	-	this study	AJ609465
ACA44	132	Ψ826, Ψ690	spliced ESTs (no apparent protein, TOP1)	1	this study	AJ609466
ACA46	135	Ψ653	hypothetical protein KIAA1007	1	this study	AJ609467
ACA50	136	Ψ38, Ψ109	hypothetical protein KIAA1007	-	this study	AJ609468
ACA60	136	Ψ1008	spliced ESTs (no apparent protein)	-	this study	AJ609469
		5.8S rRNA				
U69*	132	Ψ71	ribosomal protein L39	-	(20)	Y11163
U72	132	Ψ57	ribosomal protein L30	5	(20)	Y11167

Table 2

RNA	Length	Target	Host Gene	Additional Copies	Ref.	Acc. No
U19	200	Ψ3744, Ψ3746	U19H noncoding mRNA	-	(29)	X94290
E2	153	Ψ3731, Ψ3733	ribosomal protein SA (laminin receptor 1)	3	(48)	L07383
E3	123	Ψ4391	simlar to translation initiation factor 4A2	3	(48)	L07384
U64	134	Ψ4976	ribosomal protein S2	-	(20)	Y11158
U65	137	Ψ4374, Ψ4428	ribosomal protein L12	-	(20)	Y11159
U68	134	Ψ4394	ribosomal protein L18a	2	(20)	Y11162
ACA1	130	Ψ4442	hypothetical protein MGC5306	4	this study	AJ609425
ACA2a/b	135/137	Ψ4264, Ψ4283	hypothetical protein FLJ20436	-	this study	AJ609426/7
ACA3	130	Ψ3900, Ψ3939	ribosomal protein L27a	1	this study	AJ609428
ACA6	149	Ψ3617	ribosomal protein SA (laminin receptor 1)	-	this study	AJ609429
ACA7	139	Ψ1570, Ψ1780	ribosomal protein L32	3	this study	AJ609430
ACA9	133	Ψ1671, Ψ1770	spliced EST (no apparent protein)	3	this study	AJ609431
ACA10*	133	Ψ4492	ribosomal protein S2	1	this study	AJ609432
ACA16	134	Ψ4413	spliced ESTs (no apparent protein, TOP1)	1	this study	AJ609433
ACA17	133	Ψ4599, Ψ4660	spliced ESTs (no apparent protein, TOP2)	1	this study	AJ609434
ACA19	128	Ψ3619, Ψ3710	eukaryotic translation initiation factor 3,theta	-	this study	AJ609435
ACA21	133	Ψ4402, Ψ4471	ribosomal protein L23	-	this study	AJ609436
ACA22	134	Ψ4967	spliced ESTs (no apparent protein)	3	this study	AJ609437
ACA23	189	Ψ4332, Ψ3738	RAN binding protein 7 (importin 7)	-	this study	AJ609438
ACA27	126	Ψ3695, Ψ4523	ribosomal protein L21	3	this study	AJ609439
ACA30	129	Ψ4644	SRCAP, Snf2-related CBP activator protein	3	this study	AJ609440
ACA31*	130	Ψ3714	hypothetical protein TPT1	1	this study	AJ609441
ACA32	121	Ψ1848	hypothetical protein MGC5306	1	this study	AJ609442
ACA34	137	Ψ4270, Ψ4283	hypothetical protein FLJ20436	2	this study	AJ609443
ACA37	129	Ψ4644	methyl-CpG binding domain protein 2	3	this study	AJ609444
ACA40	127	Ψ4547	hypothetical protein MGC5306	19	this study	AJ609445
ACA43	136	Ψ4938	spliced ESTs (no apparent protein, TOP2)	2	this study	AJ609446
ACA48	135	Ψ3798	eukaryotic translation initiation factor 4A1	8	this study	AJ609447
ACA52	134	Ψ1732	ribosomal protein LP2	-	this study	AJ609448
ACA56	129	Ψ1665	dyskerin	-	this study	AJ609449
ACA58	137	Ψ3824	mitochondrial ribosomal protein L3	2	this study	AJ609450
ACA61	130	Ψ2496	spliced ESTs (no apparent protein, TOP1)	-	this study	AJ609451

Table 3

A

RNA	Length	Target	Host gene	Add. copies	Ref.	Acc. No
U85*	330	U5:Ψ46	condensin subunit 1 (chromosome condensation)	-	(25)	AF308283
U89*	270	U5:Ψ46	B-cell receptor associated protein	-	(13)	AY077739
U92	131	U2:Ψ44,Ψ34	hypothetical protein KIAA1574	-	(13)	AY077742
U93*	275	U2:Ψ54	spliced ESTs (no apparent protein)	-	(26)	AF492209
ACA12	130	U6:Ψ40	DNA polymerase alfa, catalytic subunit	-	this study	AJ609482
ACA26	129	U2:Ψ41,Ψ39	hypothetical protein KIAA0907	1	this study	AJ609483
ACA35	166	U2:Ψ89	protein phosphatase 1, regulatory (inhibitor) subunit 8	-	this study	AJ609484
ACA45	127	U2:Ψ37	spliced ESTs (no apparent protein)	2	this study	AJ609485
ACA47	187	U1:Ψ5	spliced ESTs (no apparent protein)	5	this study	AJ609486
ACA57	137	U5:Ψ43	similar to chromodomain helicase DNA binding protein 4	2	this study	AJ609487

B

RNA	Length	Host gene	Add. copies	Ref.	Acc. No
U87*	277	hypothetical protein FLJ10035	3	(13)	AY077737
U88*	266	hypothetical protein FLJ10035	1	(13)	AY077738
HBI-36	127	serotonin receptor (5HT-1C)	-	(10)	AF357423
U99	159	hypothetical protein MGC2477	-	(53)	AY349600
U100	140	hypothetical protein FLJ20516	-	(53)	AY349601
ACA11	125	Wolf-Hirschhorn syndrome candidate 1 protein	1	this study	AJ609470
ACA18	132	hypothetical protein MGC5306	3	this study	AJ609471
ACA29	140	TCP1 alpha subunit (cytopl. protein chaperone)	-	this study	AJ609472
ACA33	133	ribosomal protein S12	2	this study	AJ609473
ACA38	132	HLA-B associated transcript 2 (BAT2 protein)	-	this study	AJ609474
ACA39	136	hypothetical protein LOC128439	-	this study	AJ609475
ACA49	136	E1A binding protein p400	-	this study	AJ609476
ACA51	132	Nop56p (box C/D snoRNA-binding protein)	7	this study	AJ609477
ACA53	250	hypothetical protein SLC25A3	-	this study	AJ609478
ACA54	123	nucleosome assembly protein 1-like 4	-	this study	AJ609479
ACA55	137	poly(A)-binding protein, cytoplasmic 4	-	this study	AJ609480
ACA59	152	hypothetical protein FLJ10847	1	this study	AJ609481

Table 1. Compilation of human box H/ACA RNAs predicted to function in pseudouridylation of 18S and 5.8S rRNAs. RNAs also implicated in 28S rRNA pseudouridylation are indicated by asterisks.

Table 2. Compilation of human box H/ACA RNAs implicated in 28S rRNA pseudouridylation. Asterisks indicate guide RNAs that are predicted to function also in 18S rRNA pseudouridylation. The ACA16 RNA, under the name of U98a, has been published as an orphan H/ACA RNA (54).

Table 3. Compilation of human non-ribosomal pseudouridylation guide RNAs. (A) Box H/ACA RNAs directing pseudouridylation of spliceosomal snRNAs. (B) Box H/ACA RNAs of unknown function. Asterisks indicate composite H/ACA-C/D (U85, U87, U88 and U89) and H/ACA-H/ACA (U93) RNAs.